

Nuclear restraint is the difficult test

India's recent nuclear tests are a reminder that, despite recent progress in arms control, nuclear weapons remain a threat to humanity. Helping to prevent their use requires redoubled effort by the scientific community.

It took only a few seconds for the father of the US atomic bomb, Robert Oppenheimer, to confront the implications of his actions. "Now I am become death, the destroyer of worlds," he famously observed, when the first device was detonated above the New Mexico desert on 16 July 1945. But it took almost 20 years before the concerns of atomic scientists at the potential destructive force of their science — concerns that crystallized into a powerful movement against nuclear proliferation and atomic testing — saw their first political fruits in the partial Test Ban Treaty of 1963.

Ironically in the light of recent events, Oppenheimer was quoting from the Hindu scriptures. The Indian nuclear tests of 11 and 13 May this year have few redeeming features. In themselves, they will not give India the international status it craves, nor will they do anything to encourage the five established nuclear powers to disarm. The best that can be hoped for in the circumstances is that Pakistan's security is assured by its allies, and that peace holds until both India and Pakistan are brought into global treaties that will ultimately diminish their dependence on nuclear weapons.

This process will require the United States, Russia, China, France and Britain belatedly to adhere to their own treaty commitments and take genuinely significant steps to cut their nuclear stockpiles. Meanwhile, scientists associated either directly or indirectly with the development of the Indian bomb should take advantage of their current prestige, and press their government to behave responsibly. Initial bellicose comments last week about Indian intentions in Kashmir indicate just how difficult that will be.

Indian politicians are not unique in their adulation of nuclear weapons. In the United States, advocates of the Comprehensive Test Ban Treaty (CTBT) remain locked in battle with those whose basic instincts lean towards attaining nuclear superiority. Until two weeks ago, treaty opponents were arguing that ratification of the treaty was of no urgency, and could wait for a decade or so while US scientists figured out how to keep bombs working without testing them. Now, these opponents claim that the treaty — which India has not signed — is discredited.

The main concern of such opponents is their refusal to accept that foreign governments will adhere to treaty obligations. This point of view may have served the United States well during the Cold War. But it is insufficient for the security challenges of the coming century. On 10 June, Madeleine Albright, the US Secretary of State, will deliver a major address in Washington DC on arms control. Despite widespread nervousness about the Indian tests, these only add to the case for the treaty, and it is to be hoped that Albright will use the speech to press for its immediate ratification.

US scientists should support the administration in this endeavour, and then press for the deep cuts in nuclear weapons stockpiles advocated last June by the National Academy of Sciences (see *Nature* 387, 752; 1997). In India, nuclear scientists should capitalize on their new prestige in a bid to persuade their political masters to behave in a responsible fashion that, it is to be hoped, will now include signing up to the CTBT. Having been heroes for a day, they will discover that nuclear restraint is harder to achieve than nuclear criticality. □

Caveat investor

British Biotech's problems underline the need for scientific literacy in the stock market.

It is easy — perhaps too easy — to be cynical about the recent difficulties at British Biotech (see page 299). Those who have seen their shares in the flagship company fall to a quarter of their 1996 values in recent months — largely as a result of delays in obtaining regulatory approval for the anti-pancreatitis drug zacutex — will take little comfort from being told that they are merely learning the hard way of the volatility of biotechnology stocks. It is a lesson already familiar to US investors, who saw many similar stories in the early 1990s.

Nor is there much consolation in the departure of the company's founder and chief executive, Keith McCullagh. The move will assuage those investors who feel that McCullagh has been somewhat economical with the truth about the prospects for zacutex; it will certainly bring satisfaction to the company's former director of clinical research, Andrew Millar, who was sacked last month for giving his own, unauthorized, assessments to shareholders, which differed sharply from those of his boss. But it has done little for the credibility of Britain's biotechnology industry, on which its future investment prospects depend.

On the positive side, the whole affair shows that at least part of the

regulatory system is working well. The European Medicines Evaluation Agency seems to have been properly cautious in judging the company's clinical data to be an insufficient basis for recommending regulatory approval of zacutex. It remains to be seen how the London Stock Exchange and the US Securities and Exchange Commission handle allegations that directors sold shares before investors had been provided with bad news about the prospects for the anti-cancer drug batimastat, whose clinical trials were suspended in 1995.

If there is a silver lining, it is the way the affair has emphasized the need for greater scientific literacy in the investment community (just as, three weeks ago, media frenzy over the potential anticancer drugs angiostatin and endostatin carried a similar message for patients). With few proven products to its name, the biotech industry — like the travelling doctors of the nineteenth century — can only trade on hope. It is perhaps inevitable that the boundary between hope and hype frequently becomes blurred. Millar is to be thanked for highlighting how easily it can break down, and policing the boundary remains an important task. But in the end, *caveat investor* must remain the soundest advice. □



German technician's confession spurs check on suspect data

[MUNICH] Germany appears to be facing its second major scientific fraud scandal within a year. Following the admission by a technician at the Max Planck Institute for Plant Breeding in Cologne that she had fabricated data in at least one scientific paper, the possible systematic manipulation of important experimental results for more than six years is now being investigated.

Scientists at the institute who helped to expose the case are now repeating experiments described in more than 30 papers published in leading journals such as *Nature*, *Science*, *EMBO Journal* and *Proceedings of the National Academy of Sciences*. The publications date back to 1992.

The repeat experiments, which began in mid-April and are expected to be completed by the middle of next month, form part of a formal investigation into the affair by the Max Planck Society (MPS). Experiments in at least half a dozen papers have already been proven to be non-reproducible using specially developed tests.

The investigation will also attempt to determine how much responsibility for the length of time that the suspect data went unchallenged should be shouldered by Jeff Schell, one of the institute's four directors and head of its department of plant genetics where the experiments were carried out.

The fraud became public in March (see *Nature* 392, 111; 1998). But scientists in the department had long been suspicious about the work of a technician, Inge Czaja, as others had been unable to repeat her success with a particular assay for the enzyme adenylyl cyclase on cultured plant cells.

The role of this enzyme in hormonal signalling in plants has been controversial in the field for many years. The most recent paper from Richard Walden's group in the department, published in *Nature* last December (390, 698–701; 1997), appeared to provide definitive evidence that the enzyme exists in plants, and is part of the signalling system of the hormone auxin which promotes cell growth. The institute says it now intends to retract the paper in its previous form.

Before the investigation began, Walden



Weeding out misconduct: the Max Planck Institute for Plant Breeding is investigating some past results.

had refused to allow blind tests to be done, as requested by other group leaders in the department, to validate suspicions that the assay was being manipulated.

The fraud was exposed when colleagues gave directly to the technician samples designed to establish whether or not the assay was being manipulated. Both Czaja and Walden left the institute in February. Walden claims no direct involvement in the fraud, but admits responsibility, as group leader, for the quality of work done by his group.

Franz Weinert, an MPS vice-president and director of the Institute for Psychological Research in Munich, is chairing the investigation. He says its first main aim is to clarify how many papers have been affected by the fraud, and to what extent.

The second aim is to determine, through discussions with the ten group leaders in the department, when suspicions were first voiced, and how Schell — co-author of some of the papers being investigated — reacted to concern expressed by group leaders. Weinert says the MPS is very concerned that fabrication of data appears to have continued for so many years without being exposed.

This is the first case to be dealt with under the MPS's new rules for handling scientific misconduct which were approved last year (see *Nature* 390, 430; 1997). Under these rules, an informal and confidential internal

inquiry should be carried out within an institute when a suspicion is raised and completed within four weeks.

If grounds for suspicion are confirmed, the director must ask for a formal investigation by the society, to be carried out by a committee comprising a vice-president and three members of the MPS arbitration committee. This committee reports to the society's president, Hubert Markl, who will decide on sanctions.

In the current case, the process was initiated in the middle of April. "We are carrying out the investigation as fast as possible, while taking the time necessary to be as accurate as we can," says Weinert. The MPS will not "hush anything up", he says, and there will be "no taboos". He expects to report to Markl by the end of this month.

Last year the German scientific community was caught unprepared when two leading researchers in molecular medicine were accused of faking data over a period of at least five years in around 40 publications (see *Nature* 389, 105; 1997). In that case, too, a major concern was that the fraud had apparently been going on for a long time before exposure. Young researchers in the laboratory are said to have suspected that results were being fabricated, but were reluctant to voice suspicions outside their group.

Rules for handling such cases have now been introduced in many universities as well as in the MPS. These are designed to protect whistleblowers, as well as researchers who stand accused of misconduct.

Francesco Salamini, acting director of the Max Planck Institute of Plant Breeding, says that despite the depressed mood in Schell's department, "our scientists are doing all that can be done in such a dramatic case to find out the truth of what happened".

Alison Abbott

Nature wins publishing award

The Periodical Publishers Association (PPA) last week named *Nature* as the International Magazine of the Year in the business and professional class.

The judges cited the "heightened awareness *Nature* is attracting "throughout the world as the original source for significant scientific discoveries" and praised the new content and design introduced last year. The PPA awards are regarded as the most prestigious magazine industry awards in the UK.

Entry standards to UK science courses fall

[LONDON] Students entering physical science and engineering courses — but not life science courses — in British universities are less qualified than a decade ago, according to a survey by researchers at Brunel University. But a greater proportion of students now leave university with first-class degrees.

There are several reasons for this anomaly. The drop in university entry qualifications is almost certainly due to a fall in the standard required for mathematics, whereas the rise in first-class degrees appears to result partly from changes to assessment methods.

The survey of British science education was carried out by Alan Smithers and Pamela Robinson, director and deputy director of the Centre for Education and Employment Research at Brunel University, west London. It was funded by the Leverhulme Trust, and will be published soon.

It coincides with a controversy about university assessment techniques. Brian Dodd, a mathematics lecturer at Heriot-Watt University in Edinburgh, has angered his university by alleging in newspaper interviews that assessment methods there have been changed to reduce the student drop-out rate.

The university has denied his allegations that pass rates were reduced and the emphasis on final examinations toned down to lower the failure rate in mathematics.



The Brunel survey was based on databases kept by the Universities Statistical Record that contain details of about 600,000 students who entered and left higher education in England and Wales between 1975 and 1993. Data include subjects studied at school and grades obtained at A-level — a two-year course taken before university entrance — as well as the degree subject and class, and the first destination upon graduation.

Between 1975 and 1989, there was a doubling to 40 per cent of the proportion of physical science graduates embarking on postgraduate work, including research.

The survey reveals that average A-level

scores for students starting degree courses in mathematics dropped from 23.9 to 22.5 between 1985 and 1993. (The 'points' are the equivalent sum of grades awarded, with A=10, B=8 and so on).

In the physical sciences, scores dropped from 22.2 to 21.6, while for medicine and dentistry they increased slightly from 26 to 26.4. Only the biological sciences registered a clear increase, from 20.8 to 21.3.

But students' performance at university has improved, with more obtaining top degrees, and fewer getting low qualifications.

In 1975, 11.3 per cent of physical science graduates left universities in England and Wales with a first-class degree. By 1989 this had increased to 16.4 per cent. Similarly, in 1975, 21.3 per cent of physical science graduates obtained a third-class degree, the lowest honours degree classification. By 1989 the number had fallen to just 14.7 per cent.

Many university teachers say that the mathematics knowledge base of students embarking on degree courses in mathematics and the physical sciences has fallen consistently over the past decade. Some blame the introduction of a new examination at 16, the General Certificate for Secondary Education (GCSE), which they believe requires a lower standard of mathematics than its predecessor, the O-level examination.

Jim Matthew, professor of physics at the University of York, says the standard of mathematics and physics of students embarking on physics degree courses is worsening. Since 1979, every York physics undergraduate has sat a 'prior-knowledge' test at entry comprising 50 multiple-choice questions. The questions have remained unchanged, but Matthew says that average scores dropped "significantly" after the GCSE examination was introduced in 1990.

The reasons for the rise in the numbers of first-class degrees appear more complex. Possible factors include an increase in the numbers of women students — who are perceived to work harder than men — more individual choice in the range of subjects studied and changes to assessment methods.

Lecturers have also observed that today's students work harder than their counterparts of 10 or 15 years ago, possibly because of the much more competitive job market.

The Brunel research shows that, whereas there was a 12 per cent rise in the numbers of men studying degrees in the physical sciences between 1985 and 1993, the numbers of women increased by 52.2 per cent. The numbers of women in the biological sciences increased by 60 per cent, compared with a 20.7 per cent rise in the numbers of men.

Universities also increasingly assess students throughout the year, relying less on the traditional final examination. **Ehsan Masood**

Japanese research 'must be more flexible'

[TOKYO] Japan's research infrastructure must change to meet scientists' demands for a more open, flexible and competitive environment, according to a report by the Science and Technology Agency (STA).

A 'white paper' produced by the agency last week indicates that it is keen to improve research by adopting management styles from overseas institutions, such as Germany's Max Planck Institute and the US National Institutes of Health.

The report says that up to 70 per cent of researchers at universities and institutes are unhappy with their research environment. They complain about inflexibility in the funding system and the limited control they have over the planning of projects.

The report points out that, although the government is promoting collaboration between industry and universities, most companies prefer to do research in Europe and the United States. Most of these said the main reason was the presence of superior research organizations in these countries.

The white paper emphasizes a need to improve research management by increasing flexibility and competitiveness. It cites the Institute of Physical and Chemical Research (RIKEN) as the prime example of a

Japanese research organization that operates at an international standard.

Other science-related ministries and agencies, such as the Ministry of Education, Science, Sports and Culture (Monbusho) and the Ministry of International Trade and Industry (MITI), are trying to find a model for an 'ideal' research institute.

Their efforts are taking place amid preparations for the reorganization of nationally run research centres, part of a series of reforms aimed at streamlining Japan's government ministries and agencies.

The plan calls for the 'rationalization' of national research institutes by merging laboratories with overlapping research and turning some into independent 'agencies'. In principle, it will promote collaboration between ministries and agencies. But scientists from the organizations concerned have expressed mixed feelings about it.

For example, the merger of STA and Monbusho, scheduled for 2001, implies a possible integration of their space agencies. But researchers from the Monbusho agency are concerned that merging with the STA's applications-orientated National Space Development Agency will jeopardize their focus on basic research. **Asako Saegusa**

Panel calls for fair deal for US postdocs

[WASHINGTON] Postdoctoral fellows at US universities are treated with “an unacceptable degree of variability and instability”, according to a scathing report drawn up by the research universities themselves.

A panel of the American Association of Universities (AAU), which has been studying the problem for three years, recommends far more extensive oversight of postdoctoral education by the universities.

It calls for more reliable provision of healthcare insurance and other benefits for postdocs, and for strict time-limits that will stop people from languishing indefinitely in postdoc positions. The report also suggests that all postdocs should receive a “statement of goals, policies and responsibilities applicable to postdoctoral education” with their appointment letter.

The panel, chaired by Steven Sample, president of the University of Southern California in Los Angeles, points out that the postdoc has evolved from being an option to becoming an integral part of the training for university faculty positions.

“It is almost inconceivable today that someone would get a tenure-track post in the life sciences in a top university without having done a postdoc,” says Sample.

But the universities — including the 50 AAU members who employ the overwhelming majority of 35,000 postdocs who work in the United States, according to the report — have no policy for postdocs and exercise no oversight over their employment.

“Most institutions make little or no attempt to control the number or the quality of postdoctoral appointees on campus,” the report says. “It is common for institutions either to have no time limits on postdoctoral appointments, or regularly to ignore or waive established limits.”

The AAU report also acknowledges “the widely held perception that the postdoctoral appointment is being used as an employment holding pattern”. Privately, university officials concede that postdocs are often seen by faculty as a cheap source of laboratory assistance, without adequate attention being paid to their own education or to their career-development needs.

And although institutions claim to give postdocs full employee benefits, postdocs surveyed by the AAU cited benefits as one of their main areas of concern. The report calls for all postdocs to gain access to comprehensive health care for themselves and their families. It isn’t known how many postdocs cur-

rently lack this, according to John Vaughn, executive vice-president of the AAU.

The report finds that, of the individuals who complete postdoc appointments, a quarter move on to another postdoc post, a quarter get a tenure-track faculty position, and one-tenth get a non-tenure-track faculty post. The others leave the academic world.

It recommends that initial postdoc appointments should last no longer than two or three years, and that the total time spent in postdoctoral positions should not exceed six years. “I think it is perfectly healthy to say to people that there is a time limit on postdocs,” says Sample, adding that postdocs have to be “realistic” about the kind of job they will accept afterwards.

Neither Sample nor Vaughn could name an institution that has a statement of the goals of postdoctoral education of the type that the report suggests should be handed to postdocs with their appointment letter.

The AAU, which represents 50 leading research universities in both the public and private sectors, was founded in 1900 to coordinate the award of PhDs between universities. According to Sample, PhD education then was every bit as chaotic as postdoctoral education is today.

Colin Macilwain

Computers hope to score at the football world cup for robots

[PARIS] Scientists determined not to be left on the sidelines as France hosts football’s four-yearly world cup next month are organizing a parallel event in Paris, the annual Robot Football World Cup.

More than 60 teams of robots from major research centres worldwide will take part. The event will be the largest gathering of autonomous robots ever, and will test the prowess of the participating countries in robotics and artificial intelligence.

The matches are open to robots between 4.5 and 50 cm high, and each will last 20 minutes, with the largest robots playing five-a-side on a pitch the area of 20 table-tennis tables. The robots must play as a team autonomously; human intervention is allowed only in cases of “serious fouls or very abnormal behaviour”, such as attacking the goalposts.

Robot teams will be in permanent contact with their computer coach, however, which will analyse action on the pitch and send split-second instructions on tactics to the players.

The public competition is being organized at the Cité des Sciences et de l’Industrie in Paris from 30 June to 8 July. It is sponsored by RoboCup, a Swiss-based non-profit international association set up by researchers in artificial intelligence, and



Computer game: robots test their artificial intelligence on the pitch in last year’s world cup in Japan.

FIRA, a Korean-based organization that promotes work on cooperative autonomous robot systems.

The same venue will simultaneously host the 1998 International Conference on Multi-Agent Systems, the annual jamboree of artificial intelligence researchers.

The organizers say football poses a challenging task to robotic systems, as it tests their ability “to search, through real experimentation, the conditions for the emergence of collective intelligence”.

Such research will have applications in developing autonomous, cooperative robots for planetary exploration, military operations and domestic tasks, for example.

Leaving aside advanced cooperative algorithms, the ability of teams to ‘dribble’ — a difficult task for robots — may well be a key factor in determining the outcome of the competition itself. Winning, as in the real game, ultimately depends not on elegant play but on being able to put the ball in the back of the net.

Declan Butler

Drosophila set for fast-track sequencing

[LONDON] The humble fruitfly *Drosophila melanogaster* will be the first organism studied by the private fast-track genome-sequencing effort launched jointly earlier this month by J. Craig Venter, founder of The Institute for Genomics Research, and the laboratory equipment manufacturer Perkin-Elmer (see *Nature* 393, 101 & 201; 1998).

The main objective is to sequence the entire human genome within the next three years. But Venter has announced that he plans to sequence the 120-megabase *Drosophila* euchromatic genome as a dry run.

The project will be carried out jointly with several research teams in the United States and elsewhere already engaged in a project to sequence the *Drosophila* genome. Although this project was scheduled for completion in 2001, the approach that Venter is planning to use for his assault on the human genome should, if successful, see the task completed by next year.

"The offer [from Venter] came as a bit of a shock, as we had already drawn up our own plans," says Gerry Rubin of the University of California at Berkeley, leader of the Berkeley *Drosophila* genome project. "But it is in all our interests to see that the genome is sequenced as quickly as possible, and I feel that [this new proposal] is all going to be for the good."

Venter says that a number of potential candidate organisms were investigated for a 'proof of principle' of his sequencing strategy, which uses a 'shotgun' approach to divide the complete sequence into small fragments, each of which is sequenced several times. "We could have used the plant *Arabidopsis*, but we decided that the most useful thing we could do was *Drosophila*."

He says that one of the reasons for the choice was the large amount of work that has already been put into *Drosophila*, particularly the mapping that will allow the isolated sequences to be reassembled in the correct order. "The [existing] cDNA libraries are some of the best in the world," he says.

Sensitive to charges that his human genome efforts will be competing directly with federally funded work (see right), Venter is also keen that the *Drosophila* work should be a genuinely cooperative effort. "I hope that this will serve as a model of how we can work with others in the genome community to get sequencing done."

Rubin says he is pleased that, as part of the

collaborative agreement, Venter has agreed to make available to the international research community all the raw trace data obtained from the automatic sequencers. (One bone of contention over the human project has been the conditions under which such data will be made available to other researchers.)

The joint *Drosophila* project has also been welcomed by Francis Collins, director of the National Human Genome Research Institute, the body funding the current *Drosophila* sequencing work at Berkeley and elsewhere in the United States. "I think that we need to learn about the strategy of full genome shotgun sequencing, and what its benefits and shortcomings might be," he says. "Clearly

Drosophila would be a good test run. Its success would not necessarily predict success for the human genome; but its failure would certainly imply failure for the human genome."

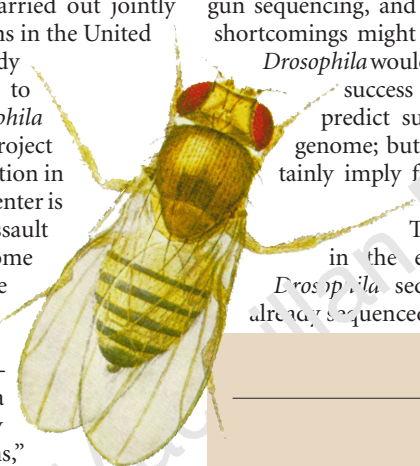
Those who are engaged in the existing international *Drosophila* sequencing, which has already sequenced about 10 per cent of

the genome, are aware that collaboration with the Venter/Perkin-Elmer initiative will involve a significant revision of some of their medium-term plans. The repercussions will also be widely felt in Europe, where laboratories in several countries, in particular the United Kingdom, Germany and Greece, are taking part in the sequencing project with support provided through the European Commission in Brussels.

Many of the researchers already engaged in the sequencing also face the prospect of having to overcome the widespread antagonism generated in the sequencing community by the aggressiveness with which news of the Venter/Perkin-Elmer initiative was announced.

A successful, cooperative *Drosophila* sequencing is seen on both sides as one way this antagonism — and the suspicion of the new initiative it has generated — could be diluted by an awareness that the sooner the sequence is completed, the sooner the more interesting work of analysing it can start. "There will still be a significant amount of work to be done," says Rubin. "The sequence is just the beginning."

David Dickson



Universities win as South Africa reverses 1997 funding trend

[CAPE TOWN] Research seems to have fared badly in South Africa's budget this year, at first glance. Nominally, it has decreased by four per cent from 1,116 million rand (US \$219 million) to R1,074 million — a decline of ten per cent in real terms after inflation is taken into account.

But science and engineering university researchers are pleased that the agency that funds research in their fields in the tertiary education sector, the Foundation for Research Development (FRD), received a budget rise of 28 per cent. This increase comes out of the science vote awarded though the Department of Arts, Culture, Science and Technology.

The allocation to universities reverses the trend in last year's budget, when the two largest councils responsible for 'in-house' research, the Council for Scientific and Industrial Research (CSIR) and the Agricultural Research Council (ARC), received increases of 19 and 11 per cent respectively (see *Nature* 386, 425; 1997).

This year both these councils had their budgets cut — by one per cent for the CSIR and 35 per cent for ARC, which has been the most severely affected.

The change has been welcomed in university circles. Renfrew Christie, dean of research at the University of the Western Cape, says it is essential that more research funds be allocated to agencies serving the higher education sector, as they are able to use them more efficiently.

The ARC has had to respond to its budget cut by terminating many projects and, in some instances, laying off staff members, according to Nico Human, its group executive for marketing and public relations.

He says the council has been successful in obtaining alternative sources of funding for some projects. But Christie counters that the ARC budget cut is both "long overdue" and not enough, describing it as an "unforgivably inefficient institution".

The FRD is using most of its increased funding to double its allocation to tertiary education institutions to buy research equipment, and to grant a 38 per cent funding increase to its programme of boosting research in 'historically disadvantaged' institutions.

The FRD also succeeded in obtaining an additional R50 million through the Department of Trade and Industry to fund its programme on technology and human resources for industry. This programme, which helps to set up partnerships between industry and universities or technical colleges, has attracted an additional R70 million from the private sector this year.

Michael Cherry

Bid to limit prices on drugs using 'public' discoveries

[WASHINGTON] The US biotechnology industry has launched a strongly worded attack on a congressional bill that seeks to restrict the prices that pharmaceutical companies can charge for products based on federally funded biomedical discoveries.

The bill would require companies that enter research collaborations with government agencies, or exclusively license rights from them, to sign a 'reasonable pricing agreement' for any products that result. The same requirement would apply to agreements between companies and government grantees, such as universities.

The bill reawakens a debate from 1995, when Harold Varmus, the director of the National Institutes of Health (NIH), removed a 'reasonable pricing' clause that had been included since 1989 in cooperative research and development agreements and exclusive licences negotiated between the NIH intramural programme and industry (see *Nature* 374, 669; 1995).

The new bill, introduced last month, is sponsored by Bernard Sanders (Independent, Vermont). It has been broadly expanded from earlier versions, first introduced in 1995, which would have applied only to agreements with industry negotiated by the NIH.

Known as the Health Care Research and Development and Taxpayer Protection Act, the new bill would prohibit federal agencies that fund biomedical research — and non-profit institutions that receive funds from them — from entering collaborations or exclusive licensing agreements unless the company involved "first agrees to a reasonable pricing agreement with the Secretary of Health and Human Services".

To arrive at a reasonable price, the bill says, the secretary should use a competitive bidding process where reasonable, but may waive the requirement if this would serve "the public interest".

Supporters of the bill — which has 14 co-sponsors, from liberal Democrats to conservative Republicans — say that it is intended to prevent drug companies from making enormous profits on products that owe their existence to government research funded by taxpayers. "There's profit, and then there's profiteering," says an aide to Fortney 'Pete' Stark (Democrat, California), a co-sponsor.

Sanders, who is the lead sponsor, calls current practice "insane". The NIH is "giving the right to drug(s) over to the private sector with no negotiations whatsoever," he says. "If the federal government has added value to the process, it is worth something; you cannot give it away."

Sanders cites the example of AZT, the first anti-AIDS drug, the development of which



Sanders: reawakening debate from 1995.

was supported by the National Cancer Institute. Burroughs-Wellcome (now Glaxo Wellcome) launched the drug in 1987 at a cost of \$10,000 per patient per year, prompting the NIH patent policy board in 1989 to require reasonable pricing clauses in agreements between the NIH intramural programme and industry.

But the bill has upset the Biotechnology Industry Organization (BIO). "It would have an adverse impact on every single relationship that the government and its grantees have with the research and development companies," says Chuck Ludlam, a BIO lobbyist, who points out that the bill targets universities as well as the NIH.

Karen Hersey, president of the Association of University Technology Managers, which represents technology transfer professionals, says that a pricing clause would cause companies to shun agreements with the NIH and other agencies, leaving valuable government research sitting on the shelf.

It would be "enormously detrimental" to the commercialization of NIH-funded research, says Hersey, who is the intellectual-property counsellor at the Massachusetts Institute of Technology. Mark Grayson, a spokesman for the Pharmaceutical Research and Manufacturers of America, agrees. Reasonable pricing clauses, he says, "reduce the reward for collaborating and in the end they stifle innovation."

Although Congress-watchers predict that Sanders' bill will not travel far in a Republican-dominated Congress, it has generated considerable attention. For example, it was featured in the *Boston Globe* newspaper last month and has been covered on the NBC television programme *Nightly News*.

Ludlam says that BIO is taking the bill "very seriously" for two reasons. First, it has grown in popularity in Congress, where it failed to pass the House of Representatives in 1996 by only 62 votes, as an amendment to the bill funding the NIH. (It was not brought up for a vote in 1997.) Second, he says, even if the bill does not pass this year, the prospect that it might do so in the near future could deter companies from entering into research agreements with universities and government agencies.

Meredith Wadman

Indian scientists speak out against bomb

[NEW DELHI] Euphoria in India about last week's nuclear tests is gradually giving way to unrest and disappointment in parts of the scientific community as plans are developed for a movement to temper New Delhi's further nuclear initiatives.

D. P. Dasgupta, professor emeritus in the electrical engineering department of the Indian Institute of Science in Bangalore, says he is "sticking my neck out" to get such a movement started. He is receiving support from physicist Sanjay Biswas and other faculty members at the institute and, he hopes, elsewhere.

A nationwide signature campaign protesting against India's development of nuclear weapons has been launched in preparation for a conference of what Dasgupta describes as "right-thinking scientists" on 30 May in Bangalore.

"Right now we are in a desperate minority," he says, aware that widespread support for the tests means that their wider implications may take time to sink in. But he is confident that the movement will grow.

Dasgupta says he has been flooded with calls from scientists from all over the country offering support after he and other speakers protesting against the tests were attacked by supporters of the nuclear bomb at a meeting in Bangalore last week.

"Some of us feel strongly about the nuclear 'weaponization' which we can ill afford when our priorities should be elsewhere," says Dasgupta.

"What is there to gloat about imitating a 50-year-old nuclear weapons technology when 50 per cent of our people are illiterate and require food, water and hospitals?" He says he was "much prouder when we developed a supercomputer on our own which the United States did not want to sell to us".

Meanwhile, 75 scientists from institutions run by the Department of Atomic Energy (DAE) have condemned the tests. Expressing "dismay and unhappiness at the action of the government", they say in a joint statement that the scientific achievement in conducting the tests has been exaggerated, and complain that India "has been committed to an expensive weapons programme without a national debate".

"The tests are bound to vitiate the atmosphere in the south Asian regions, triggering a nuclear weapons race, exacerbating the tensions that already exist, and making even more difficult the achievement of peaceful coexistence and cooperation amongst the nations of this region," the statement says.

"We don't see what immediate threats to national security forced this move, particularly when people's needs in terms of educa-



Lessons of history: activists protest in New Delhi against five nuclear tests.

tion, health, infrastructure and industrial development are urgent."

The authors of the statement come from the Institute of Mathematical science (Chennai, formerly Madras), Mehta Research Institute (Allahabad), Tata Institute of Fundamental Research (Mumbai, formerly Bombay), Institute of Physics (Bhubaneswar), Indira Gandhi Centre for Atomic Research in Kalpakkam — all of which are funded by the DAE — as well as from the Aeronautical Development Agency under the Defence Research Development Organization.

Other signatories are from the Indian Institute of Science, the five Indian Institutes of Technology, the S. N. Bose National Centre for Basic Sciences in Calcutta, the Centre for Theoretical Studies in Bangalore, and universities. Some expatriate Indian scientists in Germany, Brazil, Taiwan and Sweden also signed.

"No technocrat will oppose the Pokran [tests]," says Nataraja Sarma, a retired nuclear physicist from the DAE. "But the protests will increase when sanctions start to bite."

The political implications of the tests appear to have caused some rethinking among science policy-makers. M. G. K. Menon, a physicist and member of the ruling Bharatiya Janata Party, warned last week that the country will pay a heavy price if it decides to downgrade other sciences and give importance only to "strategic" science.

P. R. Chari, co-director of the Institute of Peace and Conflict Studies in New Delhi, expresses concern that the nuclear debate in India is dominated by "hawks", and that liberal voices are muted because dissent is considered unpatriotic.

"Indian nationalism has become inextricably linked with nuclear options and people who oppose them are considered traitors," says Kanti Bajpai, professor of international relations at the Jawaharlal Nehru University in New Delhi.

K.S. Jayaraman

Despair over pay sparks Russian strike

[MOSCOW] About 2,000 scientists went on strike in Vladivostok last week because of the government's non-payment of wages. The strike, by members of the Far Eastern branch of the Russian Academy of Sciences (RAS), formed part of a nationwide series of demonstrations.

The scientists, who blocked several major roads in the region, were also making political demands. They called for a change to the Russian president and cabinet, on the grounds that they had shown themselves to be unable to support Russian science.

"Our strike is an act of despair," said Georgyi Elyakov, a vice-president of the RAS and president of its local branch. "We have no other way to draw attention to the problems of science in the region. Previously this was one of the best scientifically productive parts of the country, with its own scientific fleet, perfectly equipped laboratories and well-financed expeditions. Now all this is gone."

The day after the strike, more than 100,000 people — including 12,000 in Moscow alone — took part in protests against education reforms proposed by the government. As well as proposing to reduce the number of lecturers by 20 per cent and

the number of students by 15 per cent, the new minister of education, Alexander Tikhonov, has drawn up a plan to give state-owned universities greater autonomy.

Meeting representatives of the academics' and scientists' trade union, Oleg Sysoev, the vice prime minister, argued that that the reforms "will draw resources from financial and other structures to the education system; at the same time the state will control the situation".

But the Russian Union of University Rectors says that it does not believe in the state's ability "to control the situation" while it is unable to support the universities financially. In 1997, education received only 2 per cent of the total state budget, and it is unlikely that this will increase in 1998.

"The other reason for our protest is the debt of 10 billion new roubles (US\$1.6 billion) that the state owes to education workers," said Nina Merkulova, an official of the central committee of the academics' and scientists' trade union. "We have not received our salaries for many months."

Russia's prime minister, Sergei Kirienko, has decided to defer a discussion of the reforms that was to have taken place at a cabinet meeting next month.

Carl Levitin

Chief to leave troubled British Biotech

[LONDON] The founder of one of Britain's first and largest biotechnology companies, British Biotech, is to resign as chief executive in an attempt to bolster shareholders' confidence in the embattled company.

The announcement of Keith McCullagh's departure was made last week in a package of measures aimed at stemming the tide of bad news that has seen the company's share price tumble (see *Nature* 392, 852; 1998).

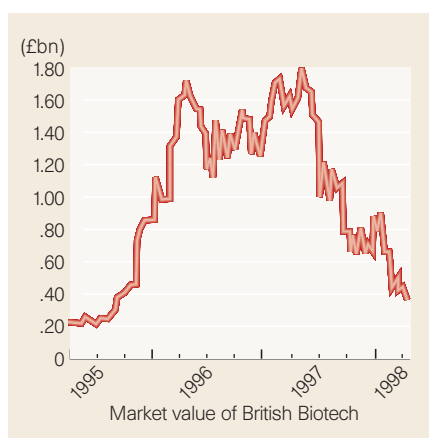
But some analysts believe that the measures may be insufficient to restore confidence in the company. They add that the episode has taught the financial community to treat assessments of the state of biotechnology company drugs trials more cautiously.

The measures announced by British Biotech, which is based in Oxford, include redundancies for 42 of the company's 300 staff. British Biotech will also seek partners in the United States to help to market its anti-cancer drug marimastat.

McCullagh will leave his post in September. His resignation was effectively forced on British Biotech following alleged insider trading, an investigation into alleged misleading statements on the progress of drugs trials and allegations that the company withheld bad news from the stock market on the progress of drugs trials.

Many of the allegations were made by Andrew Millar, the company's former director of clinical research, who was sacked for disclosing his concerns to the company's shareholders (see *Nature* 392, 746; 1998).

In a 32-page circular to shareholders, the company denies all allegations of impropriety. But some analysts feel that it raises more



questions than it tries to answer. For example, the circular says there is "no substance" to allegations that directors — including McCullagh — sold British Biotech shares before telling the London Stock Exchange of problems with the trials of an anti-cancer drug, batimastat.

The circular says that "serious adverse events" with batimastat were first reported in November 1994. The trial was eventually suspended following a board meeting on 16 February 1995 and an announcement was made the following day. A decision by the directors to sell their shares, by contrast, was taken on 15 December 1994 — before the board decided to suspend the trials.

Nick Woolf, European biotechnology analyst for the US investment bank Robertson Stephens, says the directors should not have sold shares "if they knew of questions to the viability of the batimastat study".

Equally controversial are allegations that

British Biotech withheld potentially damaging news from the stock market on problems with its anti-pancreatitis drug, zacutex.

In May 1997, the company learned that the European Medicines Evaluation Agency (EMA) would not be recommending regulatory approval for zacutex on the basis of the company's initial application because of weaknesses in the clinical data. This coincided with UK trial results showing that the drug worked best if given within 48 hours of abdominal pain to patients with severe pancreatitis.

The circular says that the company was confident that the drug would still be approved if its application was amended to recommend early treatment to acute sufferers. A press release was issued indicating the zacutex trial results, and that an application had been made for regulatory approval.

The circular says that the board decided not to make public the EMA decision, as it was preliminary, confidential and a development that was normal in the drug approval process.

But Woolf says the market should have been told of the preliminary decision. He says he appreciates the "fine line" that companies have to tread between transparency and losing investor confidence. But he says that British Biotech kept silent on the regulatory progress of zacutex for too long.

Millar, however, believes that more caution will strengthen the UK biotechnology industry, which he thinks is overvalued. "In the United States, they are much more mature, and have a regulatory system that dampens the hype," he says. **Ehsan Masood**

Germans mix support and scepticism for genetic engineering

[MUNICH] Germany's biotechnology is booming in economic terms, but genetic engineering is still struggling to win public acceptance, according to two new reports.

One report, from the international consultants Ernst and Young, lists 173 small biotechnology companies in Germany, compared with 1,300 in the United States and 200 in the United Kingdom. In addition, 269 medium-sized companies have major activities in biotechnology.

The report says that as a result of strong political support in recent years — including a closely managed university-industry competition, Bioregio, organized by the federal government (see *Nature* 379, 759; 1996) — investment is now strong in Germany.

The confidence of German investors is high, despite the fact that many companies are continuing to register heavy operating losses owing to high research costs and, as

yet, few marketable products. In contrast, however, most Germans remain deeply suspicious of 'gene technology' — as genetic engineering is known in Germany — and the scientists who use it.

The second report, from a survey coordinated by the independent Stuttgart-based Centre of Technology Assessment (CTA) and financed by the federal research ministry, shows that three-quarters of those interviewed in 1997 believe that Germany's laws on gene technology are inadequate.

The survey, whose results were presented last week at a meeting in Bonn on public acceptance of gene technology, reveals that most of those questioned believe that the safety of genetic engineering cannot be guaranteed by laws.

Only three per cent of interviewees said they trusted genetic experts. Most believe such experts are required to reflect the opinions of their employers. Even those

broadly in favour of gene technology have little confidence in the honesty and independence of genetics experts, according to the survey.

The CTA study also concluded that Germans differentiate "substantially" between gene technology and other advanced technologies, such as computing, telecommunications or space research. The latter are widely trusted to bring positive effects to society in the future. In contrast, expectations of gene technology are predominantly negative.

The report confirms the widely held notion that fears of the possible abuse of genetics, rooted in experience of the Nazi era, lie at the root of these negative attitudes. "Fears are fuelled by such ideas as eugenics, social selection, changes in the attitude towards handicapped people — even the total reinterpretation of existence," say its authors. **Quirin Schiermeier**

Nobel laureates ask Congress to boost budget for all sciences

[WASHINGTON] A group of Nobel laureates last week urged the US Congress to extend its proposed generous budget increases for the National Institutes of Health (NIH) to the other sciences. They argued that broad support is essential to the success of US science in the coming century.

The comments came at a hearing of the House Appropriations Committee subcommittee that funds the NIH. David Baltimore, the president of the California Institute of Technology, told the subcommittee that, while it is considering doubling the NIH budget over the coming years, "it is crucial that we expand research opportunities across the board in science, because the twenty-first century is going to see a cohesion of the sciences and a disappearance of their borders".

Steven Chu, a physicist at Stanford University, said that, although a doubling of the NIH budget is to be welcomed, "the time is also ripe for large increases in the other sciences". Chu won the Nobel physics prize in 1997 for the laser cooling and trapping of atoms.

Brazil ends nuclear deal with India in wake of tests

[NEW DELHI] Brazil has protested against the nuclear tests conducted by India in the second week of May by breaking off a nuclear energy cooperation agreement it had signed with the country. A letter sent to the Indian government by the Brazilian government expressed its "profound consternation at the action undertaken by India which jeopardizes the regime of non-proliferation of nuclear weapons, negatively affects the objectives of nuclear disarmament and puts at risk international peace and security".

NASA names partners for astrobiology work

[WASHINGTON] The first 11 partners have been chosen for the US space agency's new Astrobiology Institute, an experiment in 'virtual' scientific collaboration (see *Nature* 391, 109; 1998). Members of NASA's institute, linked by high-speed Internet connections, will work on research topics ranging from planetary formation and the origin of life in hydrothermal systems to the seeking of 'biomarkers' for identifying life on other planets.

The partners, who are expected to begin work by the end of June, include five universities (Harvard, University of California at Los Angeles, Colorado, Arizona State and Pennsylvania State), three research

institutions (Carnegie Institution, Scripps Research Institute and Woods Hole Marine Biological Laboratory), and three NASA centres (Johnson, Ames and the Jet Propulsion Laboratory). Projected funding for the institute is \$9 million next year and \$20 million in 2000.

Champion of Japanese science turns to politics



[TOKYO] Akito Arima (pictured), president of Japan's Institute of Physical and Chemical Research (RIKEN) and one of the most outspoken supporters of science in Japan, resigned last week to turn to politics. Arima surprised many when

he declared his intention to run for the next upper-house election, scheduled for July, as a candidate of the Liberal Democratic Party.

A former president of Tokyo University, and a member of numerous government panels, Arima helped to shape the 1996 Basic Science and Technology Law, designed to increase Japan's science spending by 50 per cent by 2001. He also introduced the external review system to Japanese universities. Originally a theoretical physicist, Arima is also a renowned *haiku* poet.

London museum opens up to Euro researchers

[LONDON] Britain's Natural History Museum has been awarded a £500,000 (US\$817,000) grant from the European Commission. The grant, which lasts for two years, is aimed at enabling researchers from the rest of Europe to access the collections of the London museum.

The funding, which comes from the commission's Training and Mobility of Researchers programme, will allow about 70 visits for up to three months each. The award also provides access to collections at the Royal Botanic Gardens at Kew, the Chelsea Physic Garden and the Linnean Society, which are also in London.

Kennedy to chair UK's council on bioethics

[LONDON] Ian Kennedy, professor of health law, ethics and policy at University College London, is to be the new chairman of Britain's Nuffield Council on Bioethics. One of the founders of the Nuffield council, Kennedy set up the centre of medical law and ethics at King's College, London in 1978.

Kennedy is also a past chairman of the UK government's advisory group on xenotransplantation.

UK food laboratories are on the move

[LONDON] The British government has confirmed its controversial decision to relocate its main food research laboratory next year from Norwich to a new complex in York (see *Nature* 392, 319; 1998). In a related development, the government's separate Institute of Food Research is to move from Reading to Norwich.

Charles Clarke, the Labour member of parliament for Norwich (South), described the decision as a "shoddy" one, which would damage food research. But his parliamentary colleague Hugh Bayley (Labour, York) said the decision was "right for British science".

Jack Cunningham, the agriculture minister, confirmed that the decision had been taken to fill empty laboratory space at the York complex.

Weapons work at US power plants ruled out

[WASHINGTON] US government plans to use civilian power plants as a source of tritium for nuclear weapons are in doubt after the House of Representatives voted overwhelmingly to prohibit the option. On a voice vote, the House passed an amendment put forward by Edward Markey (Democrat, Massachusetts) and Lindsey Graham (Republican, South Carolina) that would prohibit the use of civilian nuclear plants for tritium production.

The vote is seen by Markey as a means for the United States to stem proliferation by setting a good example to other countries. But Graham and others backed it because they want the Department of Energy to build a multi-billion dollar particle accelerator at Savannah River, South Carolina, creating jobs as well as tritium.

Stargazing company brought down to Earth

[WASHINGTON] A US company that claimed to sell the right to name stars after people has been charged with deceptive trading by the US Department of Consumer Affairs. The Illinois-based 'International Star Registry' faces a fine of at least \$3,500 for charging up to \$100 for naming stars according to the wishes of its customers.

The International Astronomical Union is the only recognized star-naming organization. And it does not sell names. Jules Polonetsky, the US consumer affairs commissioner, said that the star names were "nothing more than a listing in the company's own book". He added: "Customers who want their own star are better off saving their money by walking into Central Park, pointing to the sky, and naming it themselves."

Bioprospecting for drugs

Sir — In an article on bioprospecting¹, one of us (D. N.) was quoted to the effect that “only one sample in 250,000 will directly yield a commercial drug, although many more samples may serve as useful leads for modification through combinatorial chemistry”. It was not made clear that this estimate is based on data from the antibiotics industry^{2,3} and refers specifically to drug discovery from microbial sources. These data should not be translated directly to other natural sources, such as plants and marine organisms.

With estimates of the number of terrestrial plant species ranging from 250,000 to 400,000, the implication would be that only one or two commercial anti-cancer drugs might be developed from this source. This is an erroneous conclusion.

Indeed, the experience of the US National Cancer Institute (NCI) with the screening of plants between 1960 and 1982 indicates that some 35,000 samples (leaves, bark, roots and so on), derived from an estimated 12,000 species collected mainly from temperate regions, yielded two highly significant anti-cancer drugs, paclitaxel (Taxol) and camptothecin (which has been converted so far into two commercial products, topotecan and irinotecan), and

homoharringtonine, which shows significant clinical efficacy against several leukaemias. This gives a success rate of one in about 4,000 species tested for activity in one disease category. A substantial improvement in the ‘hit’ rate can be expected with expanded screening.

We wish to emphasize that the NCI is continuing to explore the potential of nature as a source of novel pharmacophores in collaboration with the extramural scientific community. Where feasible, collection contracts are being replaced by direct collaboration with qualified organizations in the source countries, particularly in the investigation of plants.

Gordon M. Cragg

David J. Newman

Stringer Sue Yang

Natural Products Branch,
Division of Cancer Treatment and Diagnosis,
National Cancer Institute,
PO Box B,
Frederick,
Maryland 21702-1201, USA
e-mail: cragg@dtmcp2.ncifcrf.gov

1. Macilwain, C. *Nature* **392**, 535–540 (1998).

2. Omura, S. in *The Future of Antibiotherapy and Antibiotics Research* (eds Bost, P. E. et al.) 389–405 (Academic, New York, 1981).

3. Rake, J. B. et al. *J. Antibiotics* **39**, 58–67 (1986).

No laughing matter

Sir — Andrew Birch’s pun (*Nature* **392**, 745; 1998) on gas and bubble chambers trivializes genocide in much the same spirit as the revisionists who are the subject of the accompanying article. “Revisionism” appears there euphemistically for the support of Nazi political aims and their contemporary offshoots. Birch’s belief that gas chambers are now worth a thigh-slapper indicates irresponsible ignorance or disregard of recent history.

I hope something will be done to prevent further cartoons in *Nature* joking about the other methods of the Holocaust to degrade and destroy its victims, for example beating, shooting, starvation and so on. Reflection by some on the historical role of science, especially in this century, has led to attempts to humanize the relation between science and society. Birch’s cartoon does not add to the discussion.

Michael G. Vicker

Biology/Chemistry,
University of Bremen,
D-28359 Bremen,
Germany
e-mail: vicker@uni-bremen.de

Andrew Birch replies: Far from trivializing genocide, the object of my cartoon was to ridicule the revisionists and the French National Front, and I feel that in that context the pun was justified. Just because a subject is of the utmost seriousness surely does not mean that one cannot use humour to make a political point. I obviously regret giving offence to Michael Vicker, but I hope the revisionists were more offended.

Andrew Birch

Calle Gigantes 19,
29008 Malaga,
Spain

No convergence for references

Sir — Here are a few examples of how the reference to a hypothetical paper entitled “Europe and Euro” would appear in different scientific journals:

Rossi, T. A., Bianchi, B. & Neri, C. Europe and Euro. *J. Biol. Chem.* **400**, 500–512 (1998). (*Nature*)
 Rossi TA, Bianchi B, Neri C. 1998. Europe and Euro. *J. Biol. Chem.* **400**, 500–12. (*J. Exp. Botany*)
 Rossi TA, Bianchi B, Neri C (1998) Europe and Euro. *J Biol Chem* **400**: 500–512 (*Planta*)
 Rossi TA, Bianchi B, Neri C. 1998. Europe and Euro. *J Biol Chem* **400**:500–512. (*RNA*)
 Rossi, T. A., Bianchi, B. and Neri, C., Europe and Euro. *J. Biol. Chem.* 1998. **400**: 500–512. (*Eur. J. Immunol.*)
 Rossi, T.A., Bianchi, B. and Neri, C. (1998) Europe and Euro. *J. Biol. Chem.*, **400**, 500–512. (*Embo J.*)
 Rossi, T. A., Bianchi, B. and Neri, C. (1998) Europe and Euro. *Journal of Biological Chemistry* **400**, 500–512. (*Toxicon*)
 Rossi, T.A., B. Bianchi, and C. Neri. 1998. Europe and Euro. *J. Biol. Chem.* **400**:500–512. (*J. Exp. Med.*)
 Rossi, T. A., B. Bianchi, and C. Neri. 1998. Europe and Euro. *J. Biol. Chem.* **400**: 500. (*J. Immunol.*)
 Rossi TA, Bianchi B, Neri C. Europe and Euro. *J Biol Chem* 1998, **400**, 500–512. (*Eur. J. Cancer*)
 Rossi TA, Bianchi B, Neri C. Europe and Euro. *J Biol Chem* 1998; **400**:500–12. (*Clin. Exp. Immunol.*)
 Rossi, T. A., Bianchi, B., and Neri, C. (1998) *J. Biol. Chem.* **400**, 500–512 (*J. Biol. Chem.*)
 Rossi, T. A., Bianchi, B. and Neri, C. (1998) *J. Biol. Chem.* **400**, 500–512 (*Biochem. J.*)
 Rossi, T.A., Bianchi, B. and Neri, C. (1998) *J. Biol. Chem.* **400**, 500–512. (*FEBS Lett*)
 Rossi, T.A., Bianchi, B. and Neri, C. (1998) *J. Biol. Chem.*, **400**, 500–512. (*Nucleic Acids Res.*)
 T. A. Rossi, B. Bianchi, C. Neri, J. Biol. Chem **400**, 500 (1998) (*Science*)

The list could be longer. And we have not taken into account the two alternatives for bibliography: alphabetical list of references or numbers in order of appearance. In the latter case the numbers may be 1 or 1. or [1], etc.

We propose that the year of Euro, in which 11 countries (and we hope they will soon be more) join in the use of the same currency, should also be the year in which scientific journals in Europe (and possibly throughout the world) join in the same way of handling bibliography.

Simonetta Sperti

Fiorenzo Stirpe

Maurizio Brigotti

Dipartimento di Patologia sperimentale,
Università di Bologna, Bologna, Italy
e-mail: stirpef@alma.unibo.it

Unreliable errors

Sir — The article on Emil Abderhalden and his fraudulent “defence enzymes” (*Nature* **393**, 109–111; 1998) brought back memories. When I was a graduate student at Columbia University in 1950, engaged in developing methods of peptide synthesis, I found that almost invariably I could not repeat Abderhalden’s experiments.

In one of my most frustrated moods, I asked my mentor, Dr Erwin Brand, if Abderhalden was always wrong. “No,” he said. “He is not that reliable”.

Bernard F. Erlanger

Department of Microbiology,
Columbia University,
701 West 168 Street,
New York, New York 10032, USA

Noble gases under pressure in the mantle

Minoru Ozima

New results imply that we may be in for a surprise about how argon, krypton and xenon behave in the Earth's mantle. Laboratory experiments show that these noble gases, which are used as geochemical tracers, do not act as expected under high pressure.

How did the Earth's atmosphere arise? The answer, most people think, is that it was derived from the Earth's interior or through emissions of gas from the mantle. The basic process is thought to have involved the generation of magma by partial melting of mantle material. Volatile molecules such as CO_2 and H_2O were carried with the magma and discharged from the surface to form the atmosphere, as shown in Fig. 1a and b.

In partial melting, noble gases are expelled from the solid material and dissolved into the resulting melt because of their large atomic size and chemical inertness. This characteristic behaviour is well established experimentally at ordinary to moderate pressures. Taking for granted that the result also applies at the higher pressures prevailing in the Earth's interior, mantle degassing has been assumed to be invariably accompanied by degassing of noble gases, making them excellent tracers for study of mantle degassing and the evolution of the atmosphere.

In reporting unexpected behaviour of noble gases at very high pressures, papers by Chamorro-Pérez *et al.*¹ and Jephcoat² (pages 352 and 355 of this issue) may well make us think again. If the results are indeed relevant to the mantle, they undermine the fundamental premise of noble-gas geochemistry and may force us to reconsider the generally accepted way in which the atmosphere evolved.

Chamorro-Pérez *et al.*¹ look in particular at argon in silicate melts (most of the mantle is composed of silicates), and their results should apply to all of the noble gases. They report that, in their experiments, argon solubility in such melts first increased almost linearly with pressure (in accordance with Henry's law) up to about 5 GPa, but beyond this pressure solubility abruptly decreased by more than an order of magnitude (see Fig. 2 of their paper, page 353).

Noble gases are known to be accommodated in 'holes' in silicate melts³ — or, more specifically, in the three-dimensional structure of the SiO_4 tetrahedral network⁴. Chamorro-Pérez *et al.* conclude that increas-

ing pressure makes the holes smaller, resulting in fewer sites being available to accommodate gas and its decreased solubility in melts. They further suggest that, with decrease of the hole size, the spectrum of size distribution of the holes becomes sharper; once the peak size in the spectrum becomes smaller than the radius of a molecule of

noble gas, there is an abrupt reduction in the number of accommodation sites and a sharp decrease in solubility.

If noble gases do not dissolve in melts, where do they go when the mantle undergoes partial melting? Chamorro-Pérez *et al.* do not discuss this question. However, if there is no drastic decrease in noble-gas solubility in the silicate solid at the high pressures, as Chamorro-Pérez *et al.* believe, the straightforward implication of their extraordinary experimental result is that below 150 km in the mantle (corresponding to a pressure of about 5 GPa), argon is not released from solid mantle into magma. In consequence, most of the mantle is left 'undegassed' (Fig. 1b).

This is quite contrary to what one generally assumes, for most current ideas about mantle degassing take it that the degassed mantle corresponds to the seismologically identified upper mantle (above 670 km). Do we then have to re-define the degassed mantle to a much shallower level, say down to only 150 km, and conclude that there are two types of layered mantles — a geochemically layered one and a seismologically defined one? ▶

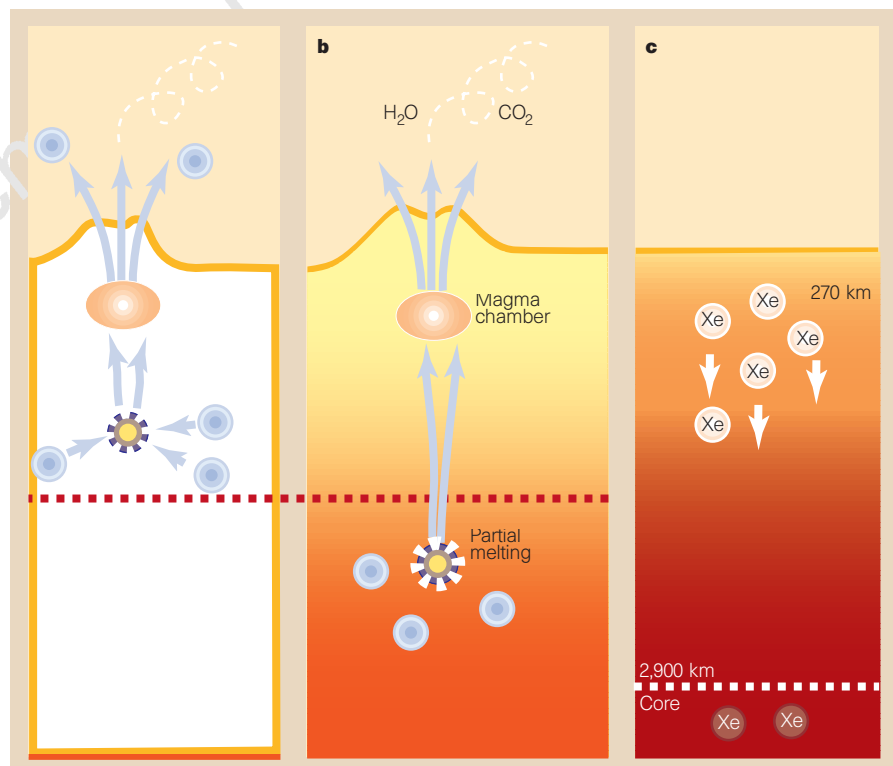


Figure 1 Scheme for how the results of Chamorro-Pérez *et al.*¹ and Jephcoat² might affect our understanding of the behaviour of noble gases in the Earth's mantle. (Also shown in parts a and b is the emission of CO_2 and H_2O from upwelling magmas, a process that is thought to have created Earth's atmosphere.) a, In partial melting of the mantle above about 150 km, argon dissolves in silicate melt (magma), is transported to the surface with the magma and leaves the mantle degassed. This view is not challenged by the new work. b, The implication of the paper by Chamorro-Pérez *et al.*¹ is that, below about 150 km, the solubility of argon in silicate melt decreases sharply. Little argon is dissolved in the magma, leaving most of the mantle essentially 'undegassed'. c, Jephcoat's results² suggest that, in most of the mantle, xenon in particular is in solid form (as are krypton and argon, but at respectively greater depths). The density of the xenon is higher than that of the surrounding mantle, and even the core; in consequence, the solid xenon particles in particular are driven deeper into the mantle. This proposed process could account for the 'missing xenon problem' — the depletion of xenon in the Earth's mantle relative to the other noble gases.

In Chamorro-Pérez and colleagues' experimental arrangement, silicate melt was surrounded by argon, which was used as the pressure medium. In these circumstances, the amount of argon in the melts was above 0.1 weight per cent. In contrast, the concentration of argon in mantle-derived materials such as volcanic rocks, which are found on the surface but should reflect argon concentration in the mantle, is generally less than 1 part per billion. Chamorro-Pérez *et al.* attribute the abrupt decrease in solubility to the availability of fewer 'holes' that can accommodate argon atoms, and for a large amount of argon dissolved in the melts this may be the case. However, argon atoms are extremely diluted in the mantle, and there may be more than enough accommodation sites even at depths where the holes are much smaller in size and number. So we need to find out whether or not the abrupt decrease of argon solubility in silicate melts at high pressures (above 5 GPa) is indeed observed for much lower concentrations of argon.

In the second paper that appears in this issue, Jephcoat² reports that the heavier noble gases (argon, krypton and xenon) become solid in the mantle, and he suggests that these solid noble gases could be stored in the deeper mantle or in the Earth's iron core without degassing. It is well established that these gases are solid at high pressures at room temperature, and recently their solid forms have been shown to melt at high temperatures (several thousand degrees Kelvin) as pressure is raised (see, for example, ref. 5).

Jephcoat compared melting temperatures at pressure, and the calculated density-pressure curves for argon, krypton and xenon at 2,500 K, with the known internal temperature and density of the Earth (Figs 2 and 3 in the paper, page 356). He found that both the melting curves and the densities for xenon and krypton are well above the estimated temperature and density of the lower mantle. Moreover, they are above about 8 GPa for xenon (corresponding to a depth of 270 km — that is, still in the upper mantle). From this it would seem that xenon, krypton and argon are mostly in solid form in the mantle. The solid xenon and krypton have much higher density than silicate throughout the whole mantle; and the calculated density of xenon even substantially exceeds that of iron in the Earth's core.

What are the implications of this work for our understanding of mantle structure? Jephcoat suggests that the solid xenon particles would segregate in the Earth's deeper regions, leaving the mantle depleted in xenon and possibly, in part, in krypton and argon as well (Fig. 1c). This would then offer a straightforward explanation for the long-standing 'missing xenon problem' — the depletion of xenon relative to other noble gases in the Earth, in comparison with its relative abundance in meteorites (see for

example my earlier News and Views article⁶ on the topic).

Up to a point this is all very satisfying. To arrive at a definitive judgement on the fate of xenon in the mantle, however, it is essential to work out the mechanism by which solid xenon particles can form and move deeper into the Earth. The mantle concentration of xenon is extremely low (less than 1 part per trillion), so one wonders how xenon particles of a size to be susceptible to gravitational segregation could be formed — at 1 part in 10^{12} , the xenon in 1 cm³ of mantle material can only form a particle a few tens of nanometres in size. As with the work of Chamorro-Pérez *et al.*, the extreme rarity of the noble gases in the mantle is the key consideration in applying

these laboratory results to the Earth.

As to applying the results to research into the evolution of Earth's atmosphere, we must wait and see. Those interested in this research topic now have a much more complicated picture to contend with. □

Minoru Ozima is in the Department of Earth and Planetary Physics, Faculty of Sciences, University of Tokyo, Tokyo 113-8653, Japan.

e-mail: diamond@gpsun01.geoph.s.u-tokyo.ac.jp

1. Chamorro-Pérez, E., Gillet, P., Jambon, A., Badro, J. & McMillan, P. *Nature* **393**, 352–355 (1998).
2. Jephcoat, A. P. *Nature* **393**, 355–358 (1998).
3. Carroll, M. L. & Stolper, E. M. *Geochim. Cosmochim. Acta* **57**, 5039–5051 (1993).
4. Shibata, T., Takahashi, E. & Matsuda, J. *Geophys. Res. Lett.* **23**, 3139–3142 (1996).
5. Jephcoat, A. P. & Besedin, S. P. *Phil. Trans. R. Soc. Lond. A* **354**, 1333–1360 (1996).
6. Ozima, M. *Nature* **321**, 813–814 (1986).

Notch signalling

A short cut to the nucleus

Julian Lewis

Our bodies have a complicated structure, but the messages that pass between cells to coordinate the construction process are simple and few in kind. A handful of basic types of signal are used repeatedly, in different tissues and organisms, to direct cells along one developmental pathway or another according to the characters of their neighbours. The signalling mechanisms have been conserved in evolution and, if we are to understand development, we need to know how they work. A report by Schroeter and colleagues¹ on page 382 of this issue, complemented by other studies appearing concurrently in *Cell*² and in *Current Biology*³, settles an old controversy and gives a remarkably neat answer to the question of how the Notch signalling mechanism — one of the key modes of cell–cell communication in animal embryos — relays messages from the surface of a cell to the nucleus.

Notch is a transmembrane protein that acts as a signal receptor and is most famous for its function in neurogenesis⁴. Neuronal cells (neurons or committed neuroblasts) normally arise as isolated individuals within a sheet of non-neuronal cells. But when Notch signalling is defective, the neighbours of neuronal cells also develop as neuronal cells, producing an excess of neurons at the expense of other cell types. Studies in *Drosophila*, *Caenorhabditis elegans* and vertebrates have shown why. Nascent neuronal cells express a Notch ligand known as Delta, which is also a transmembrane protein. By binding to Notch in the membranes of the neighbouring cells, Delta normally delivers an inhibitory signal that prevents the neighbours from developing along the neuronal pathway. Where this lateral inhibition fails, adjacent cells behave in the same way; where it oper-

ates correctly, they are forced to be different.

Signalling via Delta and Notch mediates lateral inhibition similarly in a wide range of tissues. In *Drosophila*, for example, these include the retina, the sensory bristles, the musculature, the lining of the gut, the Malpighian (excretory) tubules and the gonads. In all these cases, Delta–Notch signalling drives cell diversification, creating fine-grained patterns of contrasting cell types. Although Notch also has other functions, its most distinctive (and perhaps primordial) role is in this type of process, forcing next-door neighbours to express different sets of genes.

The link between Notch activation and gene regulation depends on the CSL protein (CBF1/Suppressor of Hairless/Lag-1). CSL can bind to specific sites in the DNA, adjacent to genes that respond to Delta–Notch signalling; it can also bind to a specific domain in the intracellular portion of Notch. But Notch lies in the cell membrane, whereas DNA lies in the nucleus; so how does the message get from the one site to the other?

Two competing answers have been proposed. According to one, it is the CSL protein that moves: activation of Notch by its ligand causes CSL to be activated and released from its binding site on Notch at the cell membrane. The CSL protein then passes into the nucleus where it regulates transcription⁵. According to the alternative view, Notch itself moves: activation by the ligand causes Notch to be cleaved, releasing the Notch intracellular domain (NICD). This then passes into the nucleus and activates transcription as part of a DNA-binding complex with CSL (refs 6–8).

Intracellular fragments of Notch produced artificially or through a mutation in the Notch gene are known to enter the

nucleus and control gene expression. Yet, so far, nobody has been able to detect accumulation of NICD in the nucleus in response to ligand binding in normal cells. Schroeter *et al.*¹, working with vertebrate cells, now show that the second view is nevertheless almost certainly correct. They have identified the site at which the transmembrane Notch-1 molecule is normally cleaved to release NICD; they have demonstrated that a mutation at this site hinders Notch-1 signalling by hindering the cleavage; and they have shown that the quantities of NICD produced in response to ligand binding are enough to regulate transcription, even though they are too small to be detected by standard immunocytochemistry.

Meanwhile, two other groups (working in *Drosophila*) have come to a similar conclusion by a different approach. Struhl and Adachi², and Lecourtis and Schweisguth³, have made flies that have either a transcriptional activator domain (Gal4 or Gal4+VP16) or a repressor domain (En or WRPW) inserted into the full-length Notch protein. Such an insert is expected to regulate transcription of an appropriate reporter gene only if the fragment of Notch that contains the insert gets into the nucleus where the reporter gene lies. The regulatory effect on the reporter is indeed seen, so long as two conditions are met — the insert has to be in the intracellular domain of Notch, and the extracellular ligand Delta has to be present. Equivalent unpublished experiments in vertebrate cells by D. Baker and D. Ish-Horowicz give similar results (see ref. 9 for meeting report). Evidently, the intracellular part of Notch is launched directly from the membrane into the nucleus in response to the external signal.

How does binding of Delta (or of the related ligand, Serrate) trigger the cleavage of Notch? Nobody knows. But, according to unpublished findings by A. Israel and colleagues⁹, the final cut that unleashes NICD is the last in a series of three, the first two cuts being made in the extracellular domain. Once the bulk of the extracellular domain has been pruned away, the final intracellular cleavage follows automatically, regardless of the presence of ligand.

This unfinished story has a tantalizing twist. The intracellular cleavage that Schroeter *et al.* have identified occurs very close to, or even in, the transmembrane domain of the Notch protein. In this it resembles the intracellular cleavage of another, more sinister transmembrane molecule — the β -amyloid precursor protein, implicated in Alzheimer's disease. The most frequent genetic cause of familial early-onset Alzheimer's disease lies in missense mutations of the presenilin-1 gene *PS1*. Presenilin-1 is a membrane protein that is thought to govern the precise location at which the β -amyloid precursor protein is

cleaved¹⁰. The same cleavage process is probably also involved in the pathogenesis of those more common forms of Alzheimer's disease that do not arise from an inherited mutation. Genetic evidence from *C. elegans* and knockout mice indicates that presenilin-1 is also a component of the Notch signalling pathway¹¹. So, as Struhl and Adachi² and other Notch-watchers have noticed, presenilin-1 is a candidate to be part of the mechanism that makes the cut that releases NICD. The same molecule that helps to endow us with a proper complement of neurons during our gestation may help to rob us of them in our dotage. □

Julian Lewis is at the Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK.

e-mail: j.lewis@icrf.icnet.uk

1. Schroeter, E. H., Kisslinger, J. A. & Kopan, R. *Nature* **393**, 382–386 (1998).
2. Struhl, G. & Adachi, A. *Cell* **93**, 649–660 (1998).
3. Lecourtis, M. & Schweisguth, F. *Curr. Biol.* (in the press).
4. Lewis, J. *Curr. Opin. Neurobiol.* **6**, 3–10 (1996).
5. Fortini, M. E. & Artavanis-Tsakonas, S. *Cell* **79**, 273–282 (1994).
6. Lieber, T., Kidd, S., Alcamo, E., Corbin, V. & Young, M. W. *Genes Dev.* **7**, 1949–1965 (1993).
7. Kopan, R., Schroeter, E. H., Nye, J. S. & Weintraub, H. *Proc. Natl Acad. Sci. USA* **93**, 1683–1688 (1996).
8. Jarriault, S. *et al.* *Nature* **377**, 355–358 (1995).
9. Bray, S. *Cell* **93**, 499–503 (1998).
10. De Strooper, B. *et al.* *Nature* **391**, 387–390 (1998).
11. Wong, P. C. *et al.* *Nature* **387**, 288–292 (1997).

Colloids

Ordering entropy

Henk N. W. Lekkerkerker and Alain Stroobants

The study of colloidal systems has revealed many puzzles. To that list can now be added further examples, as presented in the paper by Adams and colleagues that appears on page 349 of this issue¹. The authors have looked at mixtures of colloidal rods and spheres acting as 'hard' particles. Depending on the experimental conditions, the ensuing behaviour of the mixture can vary from, for instance, separation into phases in which the rods are or are not concentrated, and into alternating layers of rods and spheres.

Colloidal systems consist of small particles of one type of material in a continuous matrix of another type. They can be solid in liquid, like paint; liquid in liquid, like milk; and solid in gas, as in an aerosol. Simple yet brilliant experiments, carried out at the beginning of this century by Jean Perrin,

showed that, from a statistical point of view, a dilute suspension of colloidal particles behaves in much the same way as an ideal gas.

Perrin's view of colloids as atoms remains alive and well. Analogues of the well-known states of atomic and molecular matter (gas, liquid, crystal, liquid crystal, alloy, glass) have been observed and studied in concentrated colloidal suspensions in which the particles are interacting. Perhaps the oldest and most widely known example of a colloidal crystal, although it is not always recognized as such, is the gem opal. Opals consist of regularly packed silica particles with lattice spacings of the order of optical wavelengths. The resulting optical interference effects are the cause of opal's strikingly beautiful colours.

Why, then, study the phase behaviour of colloidal systems if it simply mimics that of atomic and molecular systems? The main

Evolutionary biology

Land-loving crabs

Like the eight other species of land crab found on Jamaica, the bromeliad crab pictured here is exceptionally well-adapted to terrestrial life. But despite their similar success in making a go of life on land, the nine species show great morphological variation, making it difficult to establish their origins by traditional means. Are they all the descendants of a single marine ancestor; did they evolve from several such ancestors; or are some more closely related to freshwater species? Elsewhere in this issue (*Nature* **393**, 363–365; 1998), Christoph D. Schubart and colleagues provide the answer.

Schubart *et al.* examined sequences from two mitochondrial genes of 22 marine and terrestrial crab species. The resulting phylogenetic trees reveal that the Jamaican terrestrial crabs originated from a single



marine forebear. Schubart *et al.* calibrated a 'molecular clock' to determine the date of separation of the land species from that ancestor as about four million years ago, relatively recently in evolutionary history. These land-loving animals then diversified rapidly, showing again that evolution isn't necessarily a lengthy business and that islands are wonderful natural laboratories for such studies.

Amanda Tromans

reason is that the interaction between colloidal particles can be tuned by modifying the surfaces of the particles and the properties of the matrix in which they are suspended. In addition, both naturally occurring and synthetic colloidal particles are available in a wide range of shapes and sizes.

Adams *et al.* have exploited both of these characteristic features of colloidal systems to produce striking examples of new and interesting phase behaviour in mixtures of rod-like virus particles and synthetic colloidal spheres. Depending on the concentrations of the colloidal rods and the colloidal spheres, and on the size ratios, Adams *et al.* observe bulk demixing into rod-rich and rod-poor phases, and microphase separation involving a variety of ordered phases. One such phase, which the authors call 'lamellar', consists of layers of rods alternating with layers of spheres. In another, metastable yet long-lived filaments of this lamellar structure coexist with either a nematic or a smectic phase. Finally, Adams *et al.* find experimental evidence for a highly ordered 'columnar' phase in which the spheres are packed in a two-dimensional array of columns orientated perpendicular to the rods.

At first sight one might think that such rich and subtle phase behaviour demands complex interactions between the particles. However, the experimental conditions are such that simple short-range repulsive interactions appear to be dominant. So, as the authors themselves say, entropy must be the driving force behind the phase transitions that they observe.

Entropy as a source of ordering might seem counterintuitive. As early as 1949, however, a seminal paper by Onsager showed unambiguously that, upon becoming concentrated, an isotropic fluid of thin, hard rods must undergo a transition to a nematic phase in which the rods have a preferred orientation. Onsager's argument was that, at sufficiently high concentration, the loss of entropy associated with the orientational ordering is more than compensated for by the gain in entropy associated with the increase in free volume in the orientationally ordered state. To understand this, one has to realize that (nearly) parallel rods take up less space than do randomly orientated rods, thus leaving more space to move in. Crystallization in an assembly of hard spheres, which was first subject to speculation by Kirkwood in 1939, and then persuasively exhibited by computer simulations in the late 1950s, might be understood on the basis of a similar argument: at sufficiently high density, the loss of entropy associated with positional ordering is more than compensated for by the gain in entropy associated with the increase in free volume.

Although the sphere case is geometrically simpler than the rod case, the arguments for entropy-driven crystallization in an assem-

bly of hard spheres are less rigorous than those for the isotropic–nematic liquid-crystal phase transition in a hard rod system. Both types of transition have been experimentally observed in colloidal model systems. Furthermore, computer simulations have provided clear evidence that more highly ordered phases in systems of rods, such as the smectic phase and the crystalline phase, can also be produced by purely repulsive interactions — which again leaves entropy as the driving force.

The work of Adams *et al.* shows that combining rods and spheres leads to much richer phase-transition phenomena than those that occur in separate systems of spheres and rods. Once again, this phase behaviour must lead to a maximization of the entropy. Ultimately this means that the loss in orientational, positional and mixing entropy associated with the bulk and microphase separations must be more than compensated for by the gain in entropy associated with the increased free volume in the ordered structures that arises as a consequence of these phase transitions.

That this balance can give rise to remarkable results was already clear from the rich variety of alloy structures observed in mixtures of hard spheres of different sizes. Bartlett *et al.*^{2,3} observed the formation of AB₂ and AB₁₃ structures in a mixture of large and small colloids A and B with a diameter ratio of 0.58. The computer simulations of Eldridge *et al.*⁴ confirm that these superlattice structures, which are highly complicated (the AB₁₃ structure has 112 particles per unit cell), result from purely entropic effects.

So the free-volume effects in mixtures of spheres of different sizes are quite subtle. Not surprisingly, then, these effects have even more dramatic consequences in mixtures of rods and spheres (as it happens, microphase separation comparable to that seen in parts b(1), b(2), c and d of Fig. 3 on page 350 has been observed in computer simulations, first in mixtures of rods of different length⁵ and later in rod–sphere mixtures⁶). These and the other novel structures observed by Adams *et al.* provide further challenges to understanding the way in which entropy produces such remarkable ordering. □

Henk N. W. Lekkerkerker and Alain Stroobants are in the Van't Hoff Laboratory for Physical and Colloidal Chemistry, Debye Institute, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
e-mail: h.n.w. lekkerkerker@chem.ruu.nl

1. Adams, M., Dogic, Z., Keller, S. L. & Fraden, S. *Nature* **393**, 349–352 (1998).
2. Bartlett, P., Ottewill, R. H. & Pusey, P. N. *J. Chem.* **93**, 1299–1312 (1990).
3. Bartlett, P., Ottewill, R. H. & Pusey, P. N. *Phys. Rev. Lett.* **68**, 3801–3804 (1992).
4. Eldridge, M. D., Madden, P. A. & Frenkel, D. *Nature* **365**, 35–37 (1993).
5. Stroobants, A. *Phys. Rev. Lett.* **69**, 2388–2391 (1992).
6. Koda, T., Numajiri, M. & Ikeda, S. *J. Phys. Soc. Jpn* **65**, 3551–3556 (1996).



100 YEARS AGO

Undoubtedly, as the editor remarks in his preface to [this] work, there has been a great desire on the part of teachers of physiology in this country to obtain a complete text-book on their subject, written in English, somewhat similar to the classical *Handbuch* of Hermann. Prof. Schäfer, with the aid of some of the best-known physiologists in Britain at the present day, has succeeded in bringing out a work which, if one may judge from the first volume, is destined to supply more or less completely the want that has been so long felt.

Text-book of Physiology, edited by E. A. Schäfer, LL.D., F.R.S.

From *Nature* 26 May 1898.

50 YEARS AGO

Physiology is a subject covering such a wide range of interests, and involving so many other branches of science, that teachers of it tend to have their own individual ways of approach. Most of them, after trying various ways of introduction to the subject, end up by considering, first, what that great teacher of physiology, Sir Michael Foster, called the 'master tissues' of the body, the muscles and the nerves.— *Introduction to Physiology*, by Prof. W. H. Newton.

★ ★ ★ ★ ★ ★ ★ ★

It is difficult to convey to this generation any adequate impression of the giant personalities that gave the Victorian era such brilliance and distinction. E. S. Beaven was one of them. Barley was his great life interest: he was the son of a barley grower, and son-in-law of a maltster whose business he entered as a youth; all his life was spent in handling barley and he knew it more intimately than any of his generation or this. But it was not only as a barley expert that he was known and respected: he was typical of his time, a sturdy, vigorous, forceful personality, outspoken, scathing in his denunciation of anything State-aided or other spoon-feeding agency; a stout believer in self-help, in some of his contemporaries, and in himself.
From *Nature* 29 May 1948.

Many more abstracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact David Plant.
e-mail: subscriptions@nature.com

Molecular neurobiology

The stress of Gulf War syndrome

Robert M. Sapolsky

Warfare leaves residues — charred remnants of what were villages, carpet-bombed moonscapes of what were rainforests, and cemeteries filled with what were 20-year-olds. These residues can include disease, such as shell shock, combat fatigue and post-traumatic stress disorder. 'Gulf War syndrome' now joins this list, a malady that is mired in controversy about its existence, its symptoms and, above all, its causes.

An important paper by Kaufer and colleagues¹ on page 373 of this issue offers a possible biological underpinning to some of the neuropsychological features of Gulf War syndrome. It also provides insights into how short-term experience becomes long-term neurochemistry, how a systemic phenomenon such as stress becomes a genomic event, and why stress disrupts cognition. Unfortunately, these findings will probably only fuel (rather than resolve) the debate over the causes of Gulf War syndrome.

The new study centres on acetylcholine. When levels of this excitatory neurotransmitter are depleted, learning and memory may be disrupted^{2,3}, indicating that it is vital to cognition. Acetylcholine is synthesized by an enzyme called choline acetyltransferase (ChAT), and it is then transported into secretory vesicles by the vesicular acetylcholine transporter (VACHT). Finally, acetylcholine is degraded by acetylcholinesterase (AChE). Thus, increased levels or activity of ChAT and VACHT increase the excitatory, 'cholinergic' tone, whereas increased levels or activity of AChE blunt it (Fig. 1).

Synaptic concentrations of acetylcholine — and, thus, neuronal excitability — can be increased by AChE inhibitors and also by stress⁴. But Kaufer *et al.*¹ now describe a hitherto unknown secondary phase of neuronal excitation in which those effects are reversed — both AChE inhibitors and stress, after causing an initial increase in excitatory cholinergic tone, lead to a long-lasting decrease. This decrease involves a prolonged increase in the expression of AChE, and a concomitant decrease in expression of ChAT and VACHT.

The authors have uncovered a plausible route by which these changes could occur (Fig. 2). By initially increasing the synaptic concentration of acetylcholine, both stress and AChE inhibitors increase the postsynaptic concentration of calcium in the cytosol. This causes the rapid induction of a transcription factor called c-Fos — a classic marker of neuronal excitation. In support of this mechanism, if the calcium is chelated c-Fos is not induced. The promoters for the genes that

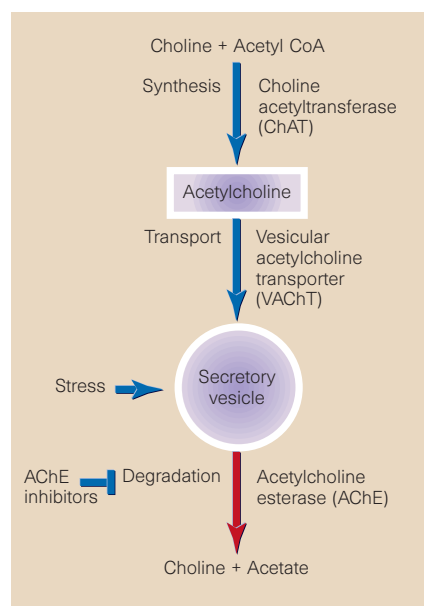


Figure 1 Compounds that affect the amount of acetylcholine in synapses. Depletion of acetylcholine can disrupt learning and memory through a decrease in neuronal excitability. Increased levels of choline acetyltransferase (ChAT) and vesicular acetylcholine transporter (VACHT) increase levels of acetylcholine, and hence neuronal excitability, whereas increased levels of acetylcholinesterase (AChE) decrease the cholinergic tone and may disrupt cognition.

encode AChE, ChAT and VACHT all contain c-Fos binding sites, and the observed changes in expression (an increase of AChE and decreases in ChAT and VACHT) occur shortly after the induction of c-Fos. So, this can be viewed as a typical feedback pathway, in which a manipulation that increases the signalling of a messenger produces a delayed compensatory decrease in such signalling. Most impressive is the duration of that compensation — in mice, Kaufer *et al.* found that cortical activity of AChE was still elevated 80 hours after the end of a swim stressor.

But there are several unresolved issues.

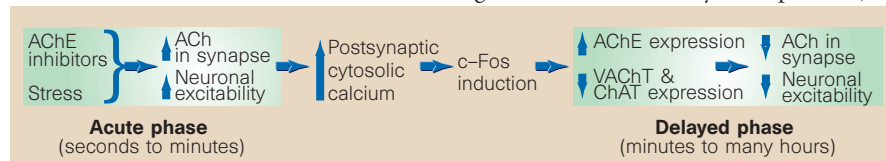


Figure 2 Factors that may contribute to the involvement of acetylcholine in Gulf War syndrome. Kaufer *et al.*¹ have discovered a pathway by which cognition could be impaired through stress and AChE inhibitors — both of which have been posited as causes of Gulf War syndrome. Stress and AChE inhibitors both lead to the production of increased levels of acetylcholine which, in turn, causes an increase in the postsynaptic concentration of calcium. The extra calcium induces expression of a transcription factor, c-Fos, which causes increased expression of AChE and decreased expression of VACHT and ChAT. In other words, this is a negative-feedback loop that leads to a long-lasting decrease in levels of acetylcholine, and, hence, decreased neuronal excitability and impaired cognition.

For example, how do such long-term transcriptional changes occur? Does c-Fos really mediate these events (something only inferred by the authors)? And why do these changes occur in cholinergic pathways in the cortex and hippocampus, but not the cerebellum? Kaufer *et al.* also found that the increased transcription of the AChE gene involved the expression of a unique splice variant, but the significance of this remains unclear. Finally, we need to determine whether the cholinergic tone is suppressed enough to impair cognition.

This study also has many implications. We know that stress can potentially disrupt cognition through the sympathetic nervous system and the secretion of glucocorticoids⁵. But the newly discovered effect does not depend on either of these factors, meaning that it constitutes a hitherto unknown route for disruption of cognition. Moreover, cholinergic pathways are damaged in Alzheimer's disease, and clinical trials with AChE inhibitors have been conducted in the hopes of boosting cholinergic tone. Unfortunately, these have generally been a failure⁶, and the delayed inhibition of cholinergic tone by AChE inhibitors might now explain those disappointing results.

The greatest significance may be for Gulf War syndrome, which is reported to involve an array of neurocognitive and mood problems, joint pain, inflammation, pulmonary complications and chronic fatigue. Two highly polarized viewpoints exist regarding its origins. Gulf War veterans argue for an organic basis arising from the toxic soup to which they were exposed. This includes biological and chemical weapons launched by the Iraqis or blown up in captured munitions depots, the toxic fumes from burning oilfields and, critically, the use of AChE inhibitors as prophylactics against nerve gas.

In contrast, the US government has doubted the legitimacy of the syndrome, suggesting that, if it exists, it is stress-related — a subspecies of post-traumatic stress disorder^{7,8}. This 'stress' argument generates great hostility among veterans — it is viewed as a means to marginalize and infer psychosomatic causes, and to side-step issues concerning exposure to toxins about which the government has already been proved (and

will probably continue) to mislead. Moreover, Gulf War veterans see this argument as a means to divert attention from the extent to which the use of AChE inhibitors was experimental and involved almost no record-keeping, as well as from the role of the United States in supplying chemical weapons to Iraq in the first place.

Kaufer and colleagues' study provides support for both arguments, although it is not clear how either variable could have effects that last years in a human, compared with just 80 hours in rodents. Whatever the resolution, some points need to be emphasized. First, Gulf War syndrome cannot be discounted simply because of its heterogeneous symptoms, given the many and varied risk factors to which troops were exposed. Second, 60 years of stress physiology undermine the view that a stress-related disease is 'merely' psychological (and thus grounds for discounting, stigmatizing or denying medical benefits or treatments). Finally, an extensive literature documents the ability of stress to worsen other

neurological insults⁹, and the present authors have published data hinting that this is the case for Gulf War syndrome as well¹⁰. We need to test whether the deleterious effects of AChE inhibitors and stress can synergize. Thus, this latest study is an important — but far from conclusive — advance for the science being conducted amid the crossfire of this contentious debate. □

Robert M. Sapolsky is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

e-mail: sapolsky@leland.stanford.edu

1. Kaufer, D., Friedman, A., Seidman, S. & Soreq, H. *Nature* **393**, 373–377 (1998).
2. Blokland, A. *Brain Res. Revs* **21**, 285–300 (1995).
3. Muir, J. *Pharm. Biochem. Behav.* **56**, 687–696 (1997).
4. Imperato, A. et al. *Brain Res.* **538**, 111–117 (1991).
5. McEwen, B. & Sapolsky, R. *Curr Opin. Neurobiol.* **5**, 205–211 (1995).
6. Knapp, M. et al. *J. Am. Med. Assoc.* **271**, 985–991 (1994).
7. Presidential Advisory Committee on Gulf War Veterans' Illnesses (US GPO, Washington DC, 1996).
8. Landrigan, P. J. *Am. Med. Assoc.* **277**, 259–261 (1997).
9. Sapolsky, R. *Stress* **1**, 1–18 (1996).
10. Friedman, A. et al. *Nature Med.* **2**, 1382–1385 (1996).

Materials science

Silk and sequence

Paul Calvert

The stiffness and strength of synthetic fibres both depend on the crystalline structure that develops during the spinning and drawing process. As the tangled mass of liquid polymer chains is sheared into a thread, small crystals form, separated by amorphous zones full of unresolved tangles. What determines the extent of crystallization? In synthetic fibres it is the degree to which the polymer is entangled, but because protein fibres such as silk are genetically encoded, it has seemed likely that the crystalline structure of silk would be controlled instead by spaced units having an amino-acid sequence conducive to β -sheet formation (crystallization) and units with sequences that inhibit such inter-chain ordering. In this way the fibre properties would result directly from the amino-acid sequence, rather than depending on the accidents of processing. But a new study by Trabbic and Yager¹ may undermine this view. It shows that the properties of silk formed naturally are the same as those of silk spun synthetically from the same proteins in solution. The study implies that the drawing process, rather than primary sequence, exerts the dominant influence on crystallinity in the fibres.

In the gland of the silkworm, *Bombyx mori*, the protein fibroin exists in a concentrated solution in a random-coil conformation. Presumably the molecules are not entangled, as in a solution of a typical synthetic polymer, but are like solutions of globular proteins and exist as separate particles. The silkworm extrudes the fluid through a

spinneret about 10 μm in diameter, and then stretches the fibre by pulling the spinneret away from a point where the thread has been attached to a support. This second elongation (drawing) step results in a stiff fibre, with crystals orientated along its axis.

For their synthetic process, Trabbic and Yager start from a solution of silk in hexafluoroisopropanol, where the protein is in a predominantly α -helical conformation. They extrude the solution through a syringe needle into methanol, where solvent exchange precipitates silk in the β -sheet form. After this they hand-draw the fibres to up to 3.5 times the original length. The natural and synthetic fibres turn out to be very

similar in degree of crystallinity, orientation and balance of helix and sheet conformations. The similarity is greatest in the case of the highest draw ratio, 3.5. It may still be that there are differences between the synthetic and natural silks, in fibre strength or stiffness or in subtle properties that help the fibre to protect the pupa from predators — but the present evidence is that a synthetically processed fibre is very like a natural fibre.

Do these similarities imply that the final fibre properties are encoded in the silk's primary structure, to become manifest no matter how the thread is prepared? On the contrary. For a high degree of crystallinity, if there are specifically irregular sections along the crystalline silk chain, they have to find one another and line up side by side, as the chains slide past one another during the spinning step — because once the fibre is solid and being drawn, there is not much opportunity for controlled rearrangement. But it does not seem very likely that the initial alignment will be the same in the natural aqueous solution as in organic solvent. It seems more likely that the drawing step slides the chains over one another to the point where there is no correlation between neighbouring chains and the structure is effectively randomized. In other words, drawing wipes the slate clean. So why have long irregular blocks separating the regular, β -sheet-forming Gly-Ala-Gly-Ala-Gly-Ser sections? The irregular regions may be important for maintaining the solution state, but have little impact on fibre structure — the draw ratio apparently determines final the properties more than the sequence.

We expected proteins to be different from synthetic polymers because the sequences are not random but are genetically specified. Polymer chemists have long regretted their inability to completely specify structure; but maybe that is not so important after all. The implication of these recent results is that the blocky structures are only going to be important for the product if the processing allows



Figure 1 Busy silkworms. To make them redundant, by synthesizing silk from scratch, we need to know how silk forms its fibre structure. The latest research hints that processing may be more important than amino-acid sequence.

HEATHER ANGEL

time for ideal chain–chain alignment to be achieved, or if some kind of post-draw annealing allows β -sheet-forming blocks to become aligned. A busy silkworm probably does not have this time to spare. The fact that the silk fibres show any crystallinity at all must instead be attributed to the predominance of crystallizable regions along the chains (forming around 70% of the total), so that many will be aligned even if the correlations between adjacent chains are random.

Supporting this picture, Thiel and Viney² show that the crystal size and crystallinity of spider silk depend on how fast the spider produces the fibre. Larger crystals form at low speeds when there is more time to match similar sequences on adjacent chains. Faster production gives smaller crystals and more tangles.

But there is evidence that the blocky primary structure does control the properties of some protein fibres. Based on sequence data, Coyne *et al.*³ describe the attachment threads of mussels as a block copolymer with elastic and collagen-like regions alternating along the chains. They regard *Bombyx* silk likewise as a blocky structure with inherently crystallizable and amorphous regions. But they may be wrong to compare silk with mussel fibre. On the synthetic side, Krejchi *et al.*⁴ have made a series of alanine–glycine polypeptides with bulky glutamate units at different intervals — (Ala–Gly)_n(Glu–Gly) where $n = 3$ to 6 — and they show that the glutamate forces the chain to fold. Thus the amino-acid sequence determines the final structure in these proteins. A synthetic fibre maker would be happy with both opinions, as crystallinity could be controlled both by processing and by random copolymerization with uncrystallizable units.

This will all really matter when we can make synthetic silk from scratch — still an unachieved goal. In the 1950s and 1960s there was a considerable effort to make artificial protein fibres^{5–7}, which foundered on the expense of amino acids as raw materials, the difficulty of polymerizing them and the unpleasant solvents needed for processing. But in the past few years, silks produced by genetically engineered *Escherichia coli* have reawakened interest in protein fibres⁸. The *E. coli* silk protein is treated to make it soluble and then spun from solution in formic acid, phosphoric acid, lithium salts or hexafluoroisopropanol. The strength and stiffness obtained with these engineered proteins is respectable — better than nylon, but not as good as Kevlar⁹.

We have been making our clothes out of protein fibres for the past 100,000 years. Perhaps we will soon have some control over the raw material. □

Paul Calvert is in the Department of Materials Science and Engineering, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: calvert@bigdog.engr.arizona.edu

1. Trabbic, K. A. & Yager, P. *Macromolecules* **31**, 462–471 (1998).
2. Thiel, B. & Viney, C. J. *Microsc.* **185**, 179–187 (1997).
3. Coyne, K., Qin, X. & Waite, J. *Science* **277**, 1830–1832 (1997).
4. Krejchi, M. *et al.* *Macromolecules* **30**, 5012–5024 (1997).
5. Noguchi, J., Tokura, S. & Nishi, N. *Angew. Makromol. Chem.* **22**, 107–131 (1972).
6. Hudson, S. M. in *Protein-Based Materials* (eds McGrath, K. P. & Kaplan, D. L.) 313–338 (Birhauser, Boston, MA, 1997).
7. Bamford, C. H. *Mater. Sci. Eng. C* **4**, 51–58 (1996).
8. Capello, J. *et al.* in *MRS Symp. Proc.: Materials Synthesis Utilizing Biological Processes* (eds Rieke, P. C., Calvert, P. D. & Alper, M.) 267–276 (Mater. Res. Soc., Pittsburgh, PA, 1990).
9. Anderson, J. P. in *Protein-Based Materials* (eds McGrath, K. P. & Kaplan, D. L.) 371–424 (Birhauser, Boston, MA, 1997).

Gene silencing

Methylation meets acetylation

Timothy H. Bestor

Promoters can sometimes be oblivious to transcription factors. For example, the ‘silent’ alleles of imprinted mammalian genes clearly have all the factors needed for their transcription, as shown by the active transcription of the other allele in the same nucleus. Such gene silencing depends on mitotic inheritance of transcriptionally repressed chromatin states. In vertebrates and flowering plants, repression is usually associated with methylation of the 5-position of cytosine residues in the underlying DNA.

Until now, nobody has known how cytosine methylation might affect the structure of chromatin. In particular, there was no link between cytosine methylation and histone acetylation. But Nan *et al.*¹ (on page 386 of this issue) and Jones *et al.*² (writing in *Nature Genetics*) show that MeCP2 — a protein that selectively binds to methylated DNA sequences — exists as a complex with histone deacetylase. The complex also contains Sin3A, a co-repressor in other deacetylation-dependent silencing processes, as well as several unidentified proteins. Both papers also show that methylation-dependent transcriptional silencing can be overcome by adding trichostatin A, which specifically inhibits histone deacetylase. The new results indicate that histone deacetylation is guided to specific chromatin domains by underlying genomic methylation patterns, and that methylation-dependent transcriptional silencing relies on histone acetylation (Fig. 1).

Histone acetylation was first tied to transcriptional activation in the mid-1960s by Allfrey and colleagues³, and in 1977 Ingram *et al.* reported⁴ that *n*-butyrate can alter cellular differentiation by inhibiting histone

deacetylation. The realization that many well-characterized transcriptional regulators exert histone deacetylase or acetyltransferase activity came much later (see ref. 5 for review). Some acetyltransferases are tightly associated with the RNA polymerase II holoenzyme, whereas other acetyltransferases and deacetylases act in concert with nuclear receptors. Moreover, the transcription of specific yeast and fruitfly promoters is abnormal in mutants that are deficient in histone acetylation or deacetylation. Although the effect on individual target genes has yet to be clarified, the different phenotypes indicate that there are many pathways of acetylation and deacetylation⁵.

How do histone acetylation and deacetylation change the structure of the nucleosome? At the moment there are two structural explanations. First, deacetylation of lysine ϵ -amino groups might allow ionic interactions between the positively charged amino-terminal histone tails and the negatively charged DNA phosphate backbone. This interaction would interfere with the binding of transcription factors to their specific DNA sequences. Second, deacetylation might lead to compaction of the chromatin by favouring interactions between adjacent nucleosomes.

The higher affinity of deacetylated histone tails for DNA favours the first explanation⁶, and the crystal structure⁷ of the nucleosome favours the second. The crystal structure — which shows the non-acetylated histone tails extending out through channels formed by the apposed minor grooves of DNA wound around the histone octamer — does not show interactions of histone tails and DNA. Of course, this could be due to the

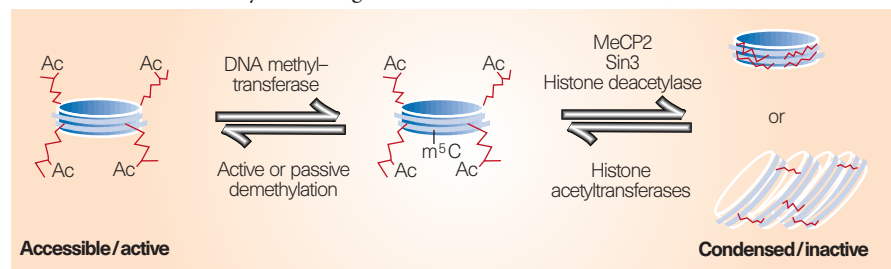


Figure 1 The effects of cytosine methylation and histone deacetylation on transcription. Transcriptional silencing in vertebrates is usually associated with the presence of 5-methylcytosine (m^5C) in the DNA. Nan *et al.*¹ and Jones *et al.*² have now discovered a link between methylation and histone deacetylation — MeCP2 (a protein that binds methylated DNA) exists in a complex with histone deacetylase.

crystallization conditions used or to crystal packing forces, and on neither side are the data decisive. It is still unclear whether the effects of histone acetylation act at the intra- or internucleosomal levels.

The experiments of Nan *et al.*¹ and Jones *et al.*² now indicate that underlying patterns of methylated cytosine residues are important in guiding histone deacetylation to specific chromatin domains. Most cytosine methylation occurs within 'silent' transposons and endogenous retroviruses⁸ — indeed, these account for at least 90% of the 5-methylcytosine in the mammalian genome, and more than 35% of the total sequence. Smaller amounts of 5-methylcytosine are found in allele-specific patterns at imprinted loci and on the inactive X chromosome of female placental mammals.

Methylated 5'CpG (cytosine–guanine) dinucleotides within promoter regions can block transcription. This can result either from major-groove clashes with regulatory factors⁹ or, more generally, through time-dependent assembly of the methylated sequences into a condensed state¹⁰. Recruitment of histone deacetylase is probably a key step in this transcriptional block. Nan *et al.*¹ found that more than 20% of the total histone deacetylase activity can be immunoprecipitated with anti-MeCP2 antibodies. Jones *et al.*² found that an even larger fraction of the histone deacetylase pool is complexed with MeCP2. These results indicate that chromatin domains could be deacetylated mainly through the methylation-dependent targeting of histone deacetylase. The ability of trichostatin A to reverse methylation-dependent transcriptional silencing suggests that most of the repression is due to histone acetylation.

Can we now be certain that chromatin acetylation patterns are templated by underlying genomic methylation patterns through MeCP2? The biochemical studies^{1,2} support this view, but some things still need to be sorted out. If all the transcriptional effects of cytosine methylation are exerted through MeCP2 and histone deacetylation, embryos deficient¹¹ in MeCP2 should show similar phenotypes to those deficient in DNA methyltransferase-1 (Dnmt1)¹². But MeCP2-defective mouse embryos do not show some of the characteristics of Dnmt1 mutant embryos, such as biallelic expression of imprinted genes or ectopic X-chromosome inactivation. Moreover, MeCP2-deficient cells differentiate normally, whereas Dnmt1 mutant embryonic stem cells undergo apoptosis when induced to differentiate. These differences could be due to additional MeCP–histone deacetylase complexes, but this has not yet been demonstrated.

Transcription of many genes in yeast and fruitflies is regulated by histone acetylation⁵, but the DNA of these organisms lacks modified bases. Marsupials manage to sustain

acetylation and transcriptional differences between the active and inactive X chromosomes, yet no allele-specific methylation has been observed at X-linked loci¹³ in marsupials. These results show that histone acetylation cannot be controlled by cytosine methylation in all cases. Furthermore, Selker has shown¹⁴ that methylation of certain loci in the ascomycete fungus *Neurospora crassa* is reduced by treatment with trichostatin A, indicating that DNA methylation patterns can respond to chromatin acetylation status. Many interesting questions remain, but the demonstration^{1,2} of a link between histone acetylation and cytosine methylation is an important step towards understanding the mechanisms of epigenetic gene silencing in vertebrates. □

Timothy H. Bestor is in the Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 West 168th

Street, New York, New York 10032, USA.

e-mail: THB12@columbia.edu

1. Nan, X. *et al.* *Nature* **393**, 386–389 (1998).
2. Jones, P. L. *et al.* *Nature Genet.* **19**, 187–191 (1998).
3. Allfrey, V., Faulkner, R. M. & Mirsky, A. E. *Proc. Natl Acad. Sci. USA* **51**, 786–794 (1964).
4. Riggs, M. G., Whittaker, R. G., Neumann, J. R. & Ingram, V. M. *Nature* **268**, 462–464 (1977).
5. Struhl, K. *Genes Dev.* **12**, 599–606 (1998).
6. Hong, L., Schroth, G. P., Matthews, H. R., Yau, P. & Bradbury, E. M. *J. Biol. Chem.* **268**, 305–314 (1993).
7. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. *Nature* **389**, 251–260 (1997).
8. Yoder, J. A., Walsh, C. P. & Bestor, T. H. *Trends Genet.* **13**, 335–340 (1997).
9. Iguchi-Arigo, S. M. & Schaffner, W. *Genes Dev.* **3**, 612–619 (1989).
10. Buschhausen, G., Wittig, B., Graessman, M. & Graessman, A. *Proc. Natl Acad. Sci. USA* **84**, 1177–1181 (1987).
11. Tate, P., Skarnes, W. & Bird, A. *Nature Genet.* **12**, 205–208 (1996).
12. Li, E., Beard, C., Forster, A. C., Bestor, T. H. & Jaenisch, R. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 297–305 (1993).
13. Wakefield, M. J., Keohane, A. M., Turner, B. M. & Graves, J. A. M. *Proc. Natl Acad. Sci. USA* **94**, 9665–9668 (1997).
14. Selker, E. U. *Proc. Natl Acad. Sci. USA* (in the press).

Aurorae

Still in the dark

Joseph E. Borovsky

After decades of research, our understanding of the aurora is disturbingly incomplete. One long-held belief, that aurorae are much more common following each peak in the 11-year solar cycle, has now been overturned by Newell and colleagues, reporting on page 342 of this issue¹. They also corroborate their earlier conclusion² that discrete aurorae occur more often above dark regions of the atmosphere than above sunlit regions. Both of these discoveries support the idea that aurorae are discharge phenomena, but they leave many questions open about the processes that drive them.

There are two main types of aurora³: diffuse and discrete. A diffuse aurora is a broad, often featureless glow, produced in the upper atmosphere by electrons and ions leaking out of the Earth's magnetosphere. Discrete aurorae, also known as auroral arcs, are bright, thin, dynamic, vertical glowing curtains produced by electrons that are somehow accelerated downward from the magnetosphere into the upper atmosphere. Narrow layers of current flow along magnetic field lines from the magnetosphere to the auroral arcs in the ionosphere. But there is no consensus on what generators power these electrical currents, or on what acceler-

ates the electrons to the high energies needed to cause the glow.

Newell *et al.*^{1,2} used satellites with particle detectors that measure the electron precipitation from the magnetosphere onto the atmosphere. Unlike visual observations of aurorae, the satellite detection works in sunlight as well as in darkness, and equally well at low as at high latitudes. They have found that discrete aurorae are more frequent above dark regions of the atmosphere than above sunlit regions, and that the occurrence of aurorae over sunlit regions is anticorrelated with the Sun's ultraviolet-output intensity.

As discrete aurorae occur where currents flow between the magnetosphere and the ionosphere, one might suspect that these light-versus-dark occurrence rates are governed by ionospheric electrical conductivity. Sunlight thickens and strengthens the ionosphere — the layer of high conductivity in the upper atmosphere (Fig. 1). But in fact, if the mechanism is based on ionospheric conductivity, the new results are counterintuitive for three reasons. First, an electrical generator in the magnetosphere that is magnetically connected to a sunlit ionosphere in one hemisphere and a dark ionosphere in the other hemisphere should be shorted out by

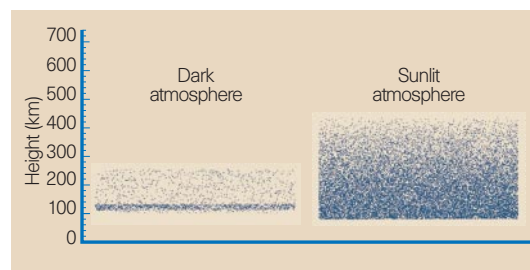


Figure 1 The layer of electrical conduction in the upper atmosphere (ionosphere) at auroral latitudes, within about 30° of each pole. The layer is greatly strengthened when sunlight shines on the upper atmosphere, which ought to mean that more aurorae happen in daylight. So why are they more common at night?

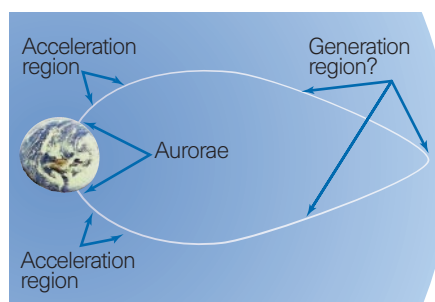


Figure 2 A magnetic field line connecting auroral regions near the Earth's North and South Poles. A current of electrons and ions, generated by an unknown mechanism far from the Earth, flows along the field lines. These charge carriers are accelerated to high energies before hitting the upper atmosphere, where they cause aurorae.

the conducting (sunlit) ionosphere, and hence the auroral currents and the discrete aurorae should favour the sunlit region. This shorting out might not happen if these generators in the magnetosphere don't reside at the Equator but rather are closer to the ionosphere^{4,5} (with these distances measured along the magnetic field lines, which carry the currents; Fig. 2). Still, it seems unlikely that such generators would only be connected to their nearest ionospheres. Second, for a voltage generator in the magnetosphere driving an auroral circuit, the more resistive (dark) the ionosphere, the less voltage is available to accelerate electrons to make aurorae^{6,7}. Third, it is hard to explain the observed slippage of fast magnetospheric flows relative to the ionospheric flow: because electrical coupling is better to a more conducting ionosphere, an aurora-producing electric field that decouples the flows is more likely to form above a conducting ionosphere than a dark ionosphere.

Newell and colleagues point out^{1,2} that their results support the ionospheric feedback model^{8,9} for auroral arcs. In this model, if the ionosphere has too low a conductivity to carry a large current from a generator, then electrons will be precipitated in the ionosphere, increasing the conductivity and so drawing more precipitation. This feedback results in a robust arc. But the shorting-out problem must still be considered if the generator has access to a sunlit and a dark ionosphere.

Another possibility, for those auroral arcs that move, is that the properties of the Alfvén waves reflecting off the ionosphere control the aurora. In such moving arcs, Alfvén waves — magnetic disturbances in a plasma — are believed to act as transmission-line transients in the magnetosphere-ionosphere circuit. The electric-field structure of a reflecting Alfvén wave depends on the conductivity of the ionosphere^{10,11}, so the manner in which electrons are accelerated and power is dissipated could be different above dark and above sunlit regions. Here,

the shorting-out problem need not arise, as fast transients are still reflected from a conducting terminus.

Another possibility is that, to complete a magnetosphere-ionosphere circuit, the current-carrying electrons in an auroral arc must have enough kinetic energy to be able to penetrate the upper atmosphere to reach the conducting layer¹². (This is very like the old notion of 'striking the arc'.) In order to reach the conducting layer from above, the electrons must be of higher energy when the atmosphere is dark than when it is sunlit (Fig. 1). Accordingly, auroral arcs, which are caused by energetic electrons, might selectively occur over dark regions, where the charge builds up until the necessary voltage is reached. But again, if the generator has access to a sunlit and a dark ionosphere, the shorting out problem comes into play.

Yet another possibility is that strong horizontal conductivity gradients can form in a dark ionosphere. This type of generator would drive currents when a flow in the magnetosphere passes through a region connected to a conductivity gradient in the ionosphere¹³.

The aurora research community has spent most of its effort and space-flight resources investigating the acceleration region of the magnetosphere (Fig. 2). But electron acceleration is only one link in the chain. The least understood, and perhaps most fundamental link is the generator. No big-picture understanding of the aurora will be obtained until space missions are launched to investigate the generator region, to determine how the magnetosphere powers aurorae and how aurorae alter the dynamics of the magnetosphere. As the critical quantities are spatial gradients — in plasma flow velocity, ion pressure and anisotropy, plasma-wave amplitude — multiple-satellite configurations will be needed. No such missions are yet planned, so until they are, aurora researchers must await enlightenment. □

Joseph E. Borovsky is in the Space and Atmospheric Sciences Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.
e-mail: jborovsky@lanl.gov

1. Newell, P. T., Meng, C.-I. & Wing, S. *Nature* **393**, 342–344 (1998).
2. Newell, P. T., Meng, C.-I. & Lyons, K. M. *Nature* **381**, 225–233 (1996).
3. Eather, R. H. *Majestic Lights: The Aurora in Science, History and the Arts* (Am. Geophys. Un., Washington DC, 1980).
4. Harrold, B. G., Goertz, C. K., Smith, R. A. & Hansen, P. J. *J. Geophys. Res.* **95**, 15039–15046 (1990).
5. Birn, J. & Hesse, M. *J. Geophys. Res.* **101**, 15345–15358 (1996).
6. Lyons, L. R. *J. Geophys. Res.* **85**, 17–24 (1980).
7. Goertz, C. K. *Space Sci. Rev.* **42**, 499–513 (1985).
8. Sato, T. & Holzer, T. E. *J. Geophys. Res.* **78**, 7314–7329 (1973).
9. Lysak, R. L. *Space Sci. Rev.* **52**, 33–87 (1990).
10. Mallinckrodt, A. J. & Carlson, C. W. *J. Geophys. Res.* **83**, 1426–1432 (1978).
11. Goertz, C. K. *Planet. Space Sci.* **32**, 1387–1392 (1984).
12. Johnstone, A. D. & Winningham, J. D. *J. Geophys. Res.* **87**, 2321–2329 (1982).
13. Rothwell, P. L., Silevich, M. B., Block, L. P. & Falthammar, C.-G. *J. Geophys. Res.* **96**, 13967–13975 (1991).

Daedalus

Current chemistry

Last week Daedalus was devising crystalline combustion catalysts with highly regular surfaces. Their product molecules were not disengaged randomly as hot gas; instead, they sprang off the ordered surface in a specific direction, carrying with them much of the energy of the reaction. Their ejection imposed a forceful directional recoil on the surface.

Thus chemical energy could be converted to mechanical energy without going through the intermediate stage of heat. Daedalus was planning aircraft, gas turbines, and so on, driven directly by reaction forces from their combustion catalyst surfaces. He now wants to generate electricity this way.

Many reacting catalytic surfaces disengage electrons; at least one (the oxidation of ethylene on a silver catalyst) does so for a form of hydrocarbon combustion. This must be a chemoelectric effect, analogous to the well-known photoelectric effect. The energy of the chemical reaction happens to be equal to the work function needed to eject an electron from the surface — so the energy is resonantly transferred to the departing electron. If the electron, coming from an ordered reaction on a regular surface, is ejected in a specific direction, the result is an electric current.

So DREADCO chemists are performing a range of hydrocarbon oxidations on regular surfaces, seeking the most copious and most ordered chemoelectric emitters. Their goal is a surface on which gas and air impinge, and from which a stream of electrons emerges. Fortunately, electrons lose little energy in interactions with gas molecules; even so, the collecting electrode must be very close to the catalyst surface — perhaps formed on it by microelectronic fabrication methods. The external circuit will return to the catalyst as a counter-electrode. Ideally, the system will burn not ethylene but butane, already used in many small catalytic-combustion devices.

DREADCO's Catabat will be almost the ultimate simple fuel cell. With no moving parts or complicated chemistry, it will take in fuel and air, and burn them to electricity. If efficient enough, it will not even get very hot. The Catabat will replace batteries, small generators, and socket power everywhere. Mobile phones will chatter longer, laptop computers will no longer fade just as the crucial file is being completed, and the global pile of discarded batteries will be replaced by a much smaller one of empty butane canisters.

David Jones

Obituary

Sir Frederick Charles Frank (1911–98)



Physicist who influenced the course of modern solid-state physics

Sir Charles Frank, who died on 5 April, joined the University of Bristol in 1946 at the invitation of Nevill Mott. Thirty years later he formally retired as head of department and H. O. Wills Professor of Physics — which, by happy coincidence, was the chair that Mott himself had occupied at the time of Frank's initial appointment as a research fellow. In his retirement address to the department, Frank defined physics as “not just concerning the Nature of Things, but concerning the Interconnectedness of all the Natures of Things”. It was this approach to physics that marked all of his work and made him one of the foremost scientists of the past 50 years.

Frank was born in Durban, South Africa, but of British parents. He was educated at schools in East Anglia, before taking degrees at Lincoln College, Oxford. He then moved to Berlin for the years 1936–38 to work with Peter Debye, the prominent physical chemist, at the Kaiser Wilhelm Physics Institute there.

In 1940, his lifelong friend and mentor, R. V. Jones, persuaded the relevant authorities to release Frank from his work at the Chemical Warfare Establishment at Porton Down, Wiltshire, so that he could join the newly formed scientific intelligence group at the Air Ministry. Jones did so, first and foremost, because Frank was an outstanding scientist, but also because Frank's time in Berlin had

given him an excellent knowledge of German and — just as significant — an understanding of the German way of doing things.

What Jones had not appreciated fully was Frank's amazing visual acuity. In January 1941, he spotted that the width of a shadow of an unusual structure near Cherbourg, in occupied France, had changed slightly in two successive reconnaissance photographs. The change was incredibly small, approximately 0.1 mm, which was essentially the resolution limit of the photograph. This acute observation led to the first identification of a German radar station. This was just the first of many major insights described in Jones's book, *Most Secret War*, published in 1978; the common thread in all of them was Frank's ability to connect apparently unconnected observations.

Frank's wide range of interests, combined with his outstanding ability, directly influenced the course of much of modern solid-state physics. His name has been associated with such diverse fields as the study of crystal growth, liquid crystals, metallurgy, polymer physics, quasicrystals, snowflakes, diamonds and geophysics. At the highest level, all of these areas involve the ability to see and describe, with the aid of rigorous mathematics, geometrical relationships in three-dimensional space. As Sir Michael Berry said in his funeral address, to see Frank's hand-waving (figuratively and literally) explanations of crystal symmetry in multidimensional space, especially when he was at the wheel of his car, was an unforgettable experience.

In spite of his reputation in solid-state physics, early in his career at Bristol Frank produced what is now regarded as a classic paper on nuclear physics. The paper, “Hypothetical Alternative Energy Sources for the ‘Second Meson’ Events” (*Nature* 160, 525–527; 1947), represented, for the first time, the idea that nuclear fusion could be catalysed by muons. As students of citation indices know, half of all published scientific papers are never cited and of the rest 80% are cited only once. A few, a very few, are cited 20 or more years after publication; today, probably 90% of papers published on muon-catalysed fusion cite that 1947 paper. Mott's Bristol was in an exciting phase of development, and although Frank was not a member of Cecil Powell's group, which was busily discovering π and μ mesons, he spotted the wider significance of these new particles and laid the foundations for the later work of Andrei Sakharov in the Soviet Union and Luis

Alvarez in the United States.

Frank's scientific style was to publish one seminal paper on a topic and then move on to new investigations. In “Curvature of Island Arcs” (*Nature* 220, 363; 1968), for example, he produced a simple geometrical explanation of the shape of island arcs. The analogy — the depression in a ping-pong ball when it is pressed in one place — was so beautiful and elegant that it just had to be right. Perhaps the most famous of his papers, “The Growth of Crystals and the Equilibrium Structure of their Surfaces” (with W. K. Burton and N. Cabrera, *Phil. Trans. R. Soc. Lond. A* 243, 299–358; 1951), laid the foundation for all that we know about crystal growth and the significance of the so-called screw dislocation as the rate-determining parameter. The impact of this contribution continues to be felt in metallurgy, materials science and the semiconductor industry. When Frank announced his findings at a conference, a member of the audience came forward and produced photographs which illustrated exactly the growth mechanism this work predicted.

Sir Charles Frank received many honours, awards and honorary degrees. His knighthood in 1977 marked his contributions to science at the national and international level. He was elected to the Royal Society in 1954 (at the comparatively young age of 43), and served as vice-president in 1967–69; he received the society's Royal Medal in 1979 and, in 1994, its highest award, the Copley Medal. The Institute of Physics recognized his work by the award of the Guthrie Medal and Prize in 1982.

In 1940 he married Maia Maita Asché, who survives him. She gave him support and encouragement throughout his long and distinguished career — Charles and Maita were a devoted couple and together did much to promote the happy atmosphere in the physics laboratory at Bristol.

Charles Frank always had a kind word for both his senior and junior colleagues, and he bore his physical discomforts with fortitude and determination (I never once heard him complain about the pain which was his constant companion in his later years). He combined intensity with humanity, even-temperedness with rigour, and presence without pomposity. We all miss him greatly.

John E. Enderby

John E. Enderby, H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK, was H. O. Wills Professor of Physics at Bristol from 1981 to 1996.
e-mail: jee@siva.bristol.ac

Max's modelling

Models of molecular structures are not just useful aids in visualization and education for our understanding of the engineering of materials, but also reflect the styles of the dominant modes of design of their period.

Martin Kemp

When Linus Pauling and Roger Hayward called their 1964 book *The Architecture of Molecules* they were resorting to an age-old style of metaphorical allusion. As knowledge of molecular structure has entered worlds that are impossible to 'see' in any normal sense, analogies with objects in our normal range of visual experience have lost none of their efficacy as aids to visualization.

One of the most spectacular acts of modelling was initiated in 1937 by Max Perutz's work on haemoglobin. As Perutz graphically recalled, "the first protein structures revealed wonderful new faces of nature". That he found it necessary to reveal these "faces" visually rather than mathematically became rapidly apparent: "I began giving a course in X-ray crystallography of biologically important molecules for students of biochemistry and other biomedical subjects. In my first lecture I introduced lattice theory, trigonometric functions, and Fourier series... but half the students failed to turn up for my second lecture.... The following year I replaced my forbidding lecture with a non-mathematical, largely pictorial introduction called 'Diffraction Without Tears'."

The potency of the visual model goes far beyond the need to sweeten the mathematical pill for students. The three-dimensional model has proved a vital tool in showing how new substances can be "engineered".

General agreement on a definitive model to 'fit' a particular X-ray diffraction pattern — as happened when James Watson and Francis Crick's model of the structure of DNA crystallized Rosalind Franklin's experi-



Watson and Crick's model of DNA, 1953.

mental data — does not solve the fundamental problems of what sort of model is appropriate. The early range extended from the spidery lattice that Watson and Crick revealed to the world in 1953 to the weighty sculptural edifice of the 1962 haemoglobin model.

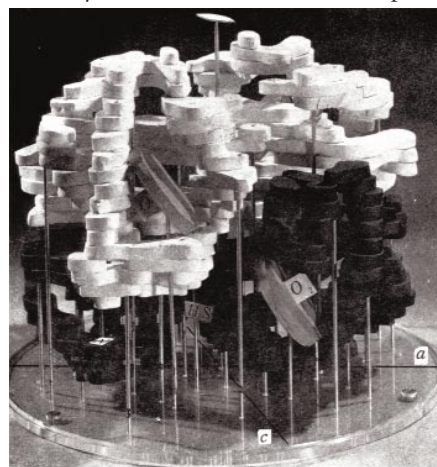
What is striking to a historian of visual images is how the models have period 'styles'. The compound of factors that make up this style — the 'look' or visual 'feel' of the object — include materials, constructional techniques, colours, textures, scales and the vocabulary of the shapes. The choices, in science no less than design, involve complex permutations of utility, technology and (often inadvertently) aesthetics.

Our sense of the period style of the Watson-Crick model — linear, wiry, openly mechanical, unadorned and rhetorically

'functional' — is framed by reference to earlier and later systems of representation. Their precarious spatial lattice stands very much within the design parameters of the 1951 Festival of Britain, the event that marked the British embrace of a modern style for a new age. In contrast, the 'Glyptic Formula' kits for modelling molecules in the nineteenth century, with their polished balls, firm rods and turned mahogany stands, exude the air of a gentleman's billiard room. And recent computer images parade the high-tech rhetoric of electronic graphics.

The haemoglobin model has its own 1960s look — assertive and futuristic like a visionary model for a concrete block of layered residences. As in any work of architecture, more is involved than mere structure. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



A. F. Cullis *et al.*'s model of the haemoglobin molecule, from Max Perutz's *Proteins and Nucleic Acids: Structure and Function*, 1962.

X-ray flare sparks quake inside Sun

Solar flares involve a release of the Sun's magnetic energy as X-radiation, particle beams and high-speed plasma flows. But we have discovered, using data from the Solar and Heliospheric Observatory (SOHO), that these flares also affect the Sun's interior, generating seismic waves similar to earthquakes. For example, a three-kilometre-high seismic wave was caused by a moderate X-ray flare that occurred on 9 July 1996 and propagated at about 50 kilometres per second to a distance 120,000 kilometres from the flare site.

This seismic response was stronger than would be expected, both from the current model of flares as giant electromagnetic explosions in the upper chromosphere and corona, and from previous attempts to detect the seismic response in ground-based data^{1,2}. The solar surface is covered by low-frequency (3 millihertz) sound waves that are stochastically excited in millions of small convective elements. These waves have been used to probe the large-scale structures of the Sun's interior³. Despite having different causes, the disturbances resulting from solar flares are fairly similar to earthquakes in many respects, for example in the impulsive localized release of energy and momentum.

The flare of 9 July was the only significant X-ray flare observed in 1996 and was of moderate size. At times of maximum solar activity, flares several times larger can be observed, which would shake the Sun more strongly.

The X-ray impulse of this flare (classified as X2.6/1B) was first detected by the Burst and Transient Source Experiment (BATSE) on board the Compton Gamma Ray Observatory at 09:07:49 UT and reached a sharp maximum at 09:09:40 (Fig. 1a). The magnetic field measurements from the Michelson Doppler Imager (MDI)⁴ instrument on SOHO show that the flare was associated with emerging flux of opposite polarity in the active region NOAA 7978 a few hours before the flare.

MDI dopplergrams reveal strong localized upward and downward mass flows during the X-ray impulse (Fig. 1b), 3–5 megametres (Mm) across. The velocity impulse was almost as sharp as the X-ray flux, but the maximum velocity was not observed until about 1 min after the X-ray maximum, at 09:11:00. This delay is consistent with theoretical predictions based on the 'thick-target' model of solar flares^{5–8}, which predicts a strong downward propagating shock following a high-energy electron beam heating the cool chromospheric 'target'. The impact of this shock causes the seismic response.

We have also detected flare ripples, circular wave packets propagating from the flare and resembling ripples from a pebble

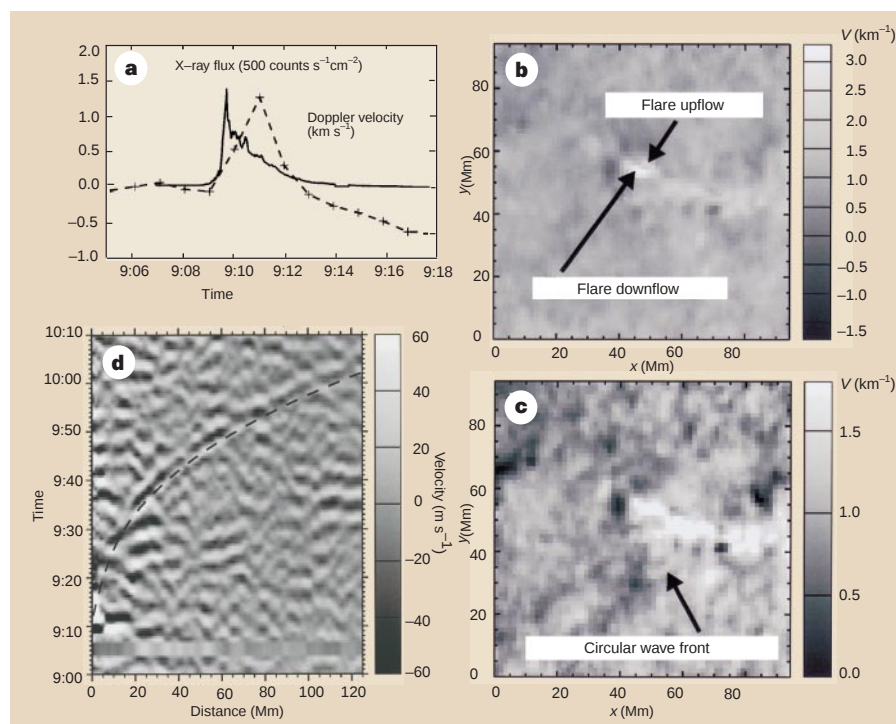


Figure 1 Characteristics of the solar flare of 9 July 1996. **a**, The X-ray flux in the energy range 25–100 keV from the BATSE flare monitor (solid curve), and the 1-minute averages of the Doppler velocity of the downflowing plasma in the flare core from the MDI (dashed curve with crosses). **b**, **c**, The MDI dopplergrams of the flare region showing the line-of-sight velocity (**b**) immediately after the X-ray pulse at 09:11 and (**c**) when the flare wave reached maximum amplitude at 09:37. Bright areas, downflows; dark areas, upflows. **d**, Flare seismogram representing the axisymmetrical components of the velocity disturbance. Dashed curve, theoretical time-distance relation for acoustic rays initiated at the flare core at 09:11; the flare signal is a black-and-white ridge around this curve. Other disturbances are solar noise.

thrown into a pond. The seismic wave was first detected at about 09:30, 20 minutes or so after the flare, at about 18 Mm from the flare. The wave was observed in a sequence of dopplergrams for about 35 minutes, after which the amplitude dropped rapidly (Fig. 1c).

We have constructed seismograms of the solar flare by remapping the Doppler images into polar coordinates centred at the point of the initial velocity impulse, and then applying a Fourier transform with respect to the azimuthal angle (Fig. 1d). A circular disturbance produced a set of ridges with a positive slope. The ridges began about 18 Mm from the flare at 09:32 and reached about 120 Mm at 10:02. The velocity of the wave packet increased from about 30 km s⁻¹ to about 100 km s⁻¹ as the wave moved from 20 to 120 Mm from the epicentre.

We have compared these observations with theoretical models of the flare seismic response, which predict similar ripples⁹. The maximum theoretical amplitude matches the observation if the shock momentum transported to the photosphere in a downward propagating shock wave is about 3×10^{22} g cm s⁻¹. This is larger than the flare

momentum typically estimated from spectral observations^{7,10}.

The dipole component of the flare wave does not have a significant signal, but the quadrupole component shows ridges at distances from 15 to 40 Mm from the flare, which presumably result from scattering on the magnetic structures of the active region.

The detection of seismic responses to solar flares opens interesting prospects for flare seismology and for determining the effects of flare processes in the Sun's interior. Substantially stronger and larger X-ray flares are observed in other stars; detection of 'starquakes' caused by such flares might be an interesting challenge for stellar astronomy.

A. G. Kosovichev

W. W. Hansen Experimental Physics Laboratory,
Stanford University, Stanford,
California 94305-4085, USA

V. V. Zharkova

Department of Physics and Astronomy,
Glasgow University, Glasgow G12 8QQ, UK

1. Haber, D. A., Toomre, J., Hill, F. & Gough, D. O. in *Proc. Symp. Seismology of the Sun and Sun-like Stars*, Tenerife, Spain, 26–30 September, ESA SP-286, 301–304 (1988).
2. Braun, D. C. & Duvall, T. L. Jr *Sol. Phys.* **129**, 83–94 (1990).
3. Kosovichev, A. G. et al. *Sol. Phys.* **170**, 43–61 (1997).
4. Scherrer, P. H. et al. *Sol. Phys.* **162**, 129–188 (1995).

5. Kostyuk, N. D. & Pikel'ner, S. B. *Sov. Astronomy* **18**, 1002–1016 (1975).
6. Kosovichev, A. G. *Bull. Crimean Astrophys. Observatory* **75**, 6–18 (1986).
7. Zarro, D. M., Canfield, R. C., Strong, K. T. & Metcalf, T. R. *Astrophys. J.* **324**, 582–589 (1988).
8. Zharkova, V. V. & Brown, J. C. in *Solar Dynamic Phenomena and Solar Wind Consequences*, Proc. 3rd SOHO Workshop, ESA SP-373, Noordwijk, 61–65 (1994).
9. Kosovichev, A. G. & Zharkova, V. V. in *Helioseismology*, Proc. 4th SOHO Workshop (eds Hoeksema, J. T., Domingo, V., Fleck, B. & Battrock, B.), ESA SP-376, Noordwijk, 341–344 (1995).
10. Canfield, R. C., Zarro, D. M., Metcalf, T. R. & Lemen, J. R. *Astrophys. J.* **348**, 333–340 (1990).

Low nitrate:phosphate ocean ratios corrected...

It has become apparent, since publication of our Letter on low nitrate:phosphate ratios in the global ocean (T. Tyrrell and C. S. Law, *Nature* **387**, 793–796; 1997; correction, **393**, 396; 1998), that the World Ocean Atlas 1994 (WOA94) database we used contains transcription errors for nutrient measurements. Discrepancies between the original data and the corresponding data sets in WOA94 have been confirmed for the eastern tropical Pacific (identified by A. Longhurst and S. Sutherland), the western North Pacific (M. Aoyama, K. Hirose and K. Ishikawa) and the Agulhas retroflection area south of South Africa (R. Schlitzer and A. Longhurst).

Removal of falsely identified low nitrate:phosphate points (LNP, $[(\text{NO}_3^-) \div (\text{PO}_4^{3-})] < 3.0$ and $[\text{PO}_4^{3-}] > 1.5 \mu\text{mol kg}^{-1}$) in these areas shows that: (1) LNP points do not occur in the open ocean away from the coast; and (2) the western and northern North Pacific does not represent a hitherto-unquantified significant site for denitrification, as we originally suggested.

However, by examining other independent and more rigidly controlled data sets and re-evaluating WOA94 data using stricter criteria, we have confirmed that LNP is a feature of the Baltic and Black Seas, the Peruvian upwelling system, the Cariaco trench, the Bering Straits, and Tomales Bay. Although this confirms our original premise that the Redfield ratio is not universally applicable, it is clear that LNP points are restricted to coastal and enclosed shelf sea regions. These observations accord with current understanding of the extent of denitrification.

We thank those named above for detecting data discrepancies, and Steve Smith, Louis Codispoti, Geoff Bailey, Sergey Konovalov, Lee Cooper, Peter McRoy, Chirk Chu, Jean Garside, Chris Garside, Gwo-Ching Gong, Alexander Bychkov and Mikael Krysell for sending data sets.

T. Tyrrell

Department of Oceanography, Southampton
Oceanography Centre, University of Southampton,
European Way, Southampton SO14 3ZH, UK

C. S. Law

Plymouth Marine Laboratory, Prospect Place,
Plymouth PL1 3DH, UK

...and database flagged

The errors in the World Ocean Atlas 1994 that were identified as a result of the publication of Tyrrell and Law's Letter (T. Tyrrell and C. S. Law, *Nature* **387**, 793–796; 1997) are, as part of our routine quality-control procedures, being corrected or flagged as suspect on this and related National Oceanographic Data Center databases.

Margarita Conkright

Ocean Climate Laboratory, OC5, National
Oceanographic Data Center, 1315 East West
Highway, Silver Spring, Maryland 20910-3282, USA

Global impact of the 1789–93 El Niño

It has been suggested that global warming has caused the El Niño/Southern Oscillation (ENSO) climatic events to become more frequent and intense. However, several ENSO events that occurred before 1880 had effects at least as intense and wide-ranging as those associated with the current event. This is the case particularly for the events in 1396 (ref. 1), 1685–88, 1789–93 and 1877–79. Here I discuss archival evidence, notably from South Asia and above all for the 1789–93 ENSO, for the strength of these historical effects.

In peninsular India, every major drought between 1526 and 1900 has been closely associated with the eastern Pacific El Niño. The 1789–93 ENSO event produced prolonged droughts, especially in South Asia, a region where the association between ENSO and the monsoon is well established². The global impact of this event was recognized in 1816 by Alexander Beatson, governor of St Helena, who suggested that the 1791 droughts in India, St Helena and Montserrat were part of a single, connected phenomenon³.

The earliest indications of the event in the tropics were meteorological observations, based on a 14-year data set started in the early 1770s by William Roxburgh at Samulcottah (Samalkot) in southern India⁴. He

recognized the exceptional nature of the droughts beginning in 1789 and believed that their severity had only been approached by those in 1685–87 (ref. 5). Those years are now known to have been characterized by 'very severe' El Niño in the eastern Pacific in 1687–88 (ref. 6).

Roxburgh recorded the continuous failure of the South Asian monsoon between 1789 and 1792, the severest failure being in 1790 (Table 1). The first major rainfall reduction was in 1789 in southern India, more than a year before similar droughts occurred in Australia, Mexico, the Atlantic islands and southern Africa. By November 1792, 600,000 deaths were attributed to the resultant droughts in the northern Madras Presidency alone, where half the population died.

The long droughts were interspersed with short periods of highly destructive rainfall. In three days at Madras in late October 1791, 25.5 inches of rain fell, "more than... has been known within the memory of man"⁷. Unseasonal severe droughts were experienced in Java, and in New South Wales, Australia, in the same year. On 5 November 1791, the governor of this colony, Arthur Phillip, reported that the normally perennial 'Tank Stream' river flowing into Sydney Harbour had been dry for "some months". It did not flow again until 1794. Phillip marks the start of the droughts in July 1790; no rain had fallen by August 1791 (ref. 8).

In Mexico, the level of Lake Pátzcuaro fell steadily between 1791 and 1793, giving rise to disputes over the ownership of the land that emerged⁹. By mid-August 1791, the desiccating effects on the islands of the Antilles were the severest since 1700, and no rain had fallen on the islands of St Vincent and Montserrat¹⁰. The drought continued on Montserrat until November 1792.

Droughts on St Helena were later than those in the Caribbean, lasting from late 1791 to mid-1794. The River Nile fell to very low levels from 1790 to 1797, as a result of reduced rainfall in the Ethiopian highlands. Evidence from the rest of Africa is scanty, but prolonged droughts in Natal and Zululand between 1789 and 1799 resulted in the *Mahlatule* famine, the severest known to have affected Southern Africa before 1862 (ref. 11).

This evidence that the 1789–93 ENSO had a strong global impact indicates that it was

Table 1 Monthly rainfall at Samulcottah, Andhra Pradesh, India: May–November 1788–92 (in inches and twelfths of an inch)¹⁵

	1788	1789	1790	1791	1792
May	15.4	1	–	4	3.6
June	7.2	6	1.8	4.1	5
July	22.3	6.10	4.9	5.6	6.4
August	12.2	21.1	3.8	–	1.8
September	8.9	1.4	4.8	3.9	7.5
October	5.9	10.1	1.5	3.3	13.11
November	6	1.3	1.2	6.4	–
Total	77.5	43.10	17.4	26.11	37.10

one of the most severe on record. An early precursor of the event may have been an unusually cold winter in western Europe in 1787–88, followed by a late and wet spring and summer drought, resulting in the crop failures preceding the French Revolution¹². This may have resulted from a weak phase in the North Atlantic Oscillation, which would fit an established correlation between inter-annual variability in the North Atlantic Oscillation and Indian summer monsoon rainfall¹³. Indeed, the Indian monsoon is an active feature of tropical circulation and monsoon failure is efficient in foreshadowing ENSO¹⁴.

As with the 1685–88 ENSO, the early stages of the 1789–93 event were observable in southern India more than a year before the El Niño effect was recorded in the Pacific basin. However, a few ENSO events, including the present event of 1997–98, coincide with a failure of the Southeast Asian monsoon rather than that of South Asia. In the case of the 1789–93 ENSO event, the monsoon failed in both regions.

So the developmental sequence of the 1789–93 ENSO is important as a basis for comparison with other, very severe, ENSO events.

Richard H. Grove

Institute of Advanced Studies, Australian National University, Canberra, ACT 0200, Australia
email: grove@coombs.anu.edu.au

1. Duff, J. C. G., *A History of the Mahrattas* Vol. 1, 50 (Murray, London, 1826).
2. Charles, C. D. *Science* **277**, 925–928 (1997).
3. Beaton, A. *Tracts on the Island of St Helena*, 10 (Black, London, 1816).
4. Roxburgh, W. *Phil. Trans. R. Soc.* **68**, 180–190 (1778); **80**, 112–113 (1790).
5. Roxburgh, W. *MS Report to the President's Council* Dated 6 February, in East India Company Boards Collections, ref. no. F/4/99 (British Library India Office Library and Records, London, 1793).
6. Quinn, W. H., Neal, V. T. & Antunez de Mayolo, S. E. *J. Geophys. Res.* **92**, 14449–14461 (1987).
7. *The Madras Courier*, 3 November, 1 (1791).
8. Nicholls, N. *Bull. Am. Meteorol. Soc.* **69**, 4–7 (1988).
9. Endfield, G. & O'Hara, S. *Environment and History* **3**, 4–22 (1997).
10. McKealy, W. *Petition on Behalf of the Council of Montserrat*. Dated 13 August, in Montserrat Legislative Assembly Proceedings (Government Archives, Crown Colony of Montserrat, British West Indies, 1791).
11. Hall, M. *Ann. Natal Museum* **702** (1976).
12. Davies, A., *History* **49**, 37–38 (1964).
13. Dugam, S. S. et al. *Theor. Appl. Meteorol.* **58**, 21–29 (1997).
14. Normand, C. Q. J. *Roy. Meteorol. Soc.* **79**, 463–473 (1953).
15. Roxburgh, W. *Letter to Sir Henry Oakley*. Dated 23 January, in East India Company Boards Collections, ref. no. F/4/99, 29 (British Library India Office Library and Records, London, 1793).

Connexin 26 gene linked to a dominant deafness

A high proportion of all cases of congenital deafness is caused by mutations in a gene coding for a gap-junction protein, connexin 26. The deafness associated with this gene, Cx26, is the autosomal recessive

form, DFNB1 (refs 1–3); its involvement in autosomal dominant forms of deafness has remained controversial⁴. Here we show that a mutation in Cx26 underlies the dominant form of deafness, DFNA3.

About 70% of prelingual hereditary deafness is non-syndromic (or isolated) sensorineural deafness⁵. Autosomal recessive forms (DFNB) make up about 85% of these cases, autosomal dominant forms (DFNA) 12–15%, and X-linked forms (DFN) 1–3%. The postlingual forms have been less well studied, but seem to be mainly sensorineural defects with a dominant mode of transmission in a high proportion of cases. The non-syndromic forms of deafness are extremely heterogeneous, with up to 100 genes being involved.

So far, at least 15 DFNB and 14 DFNA loci have been mapped to human chromosomes^{6,7}. However, mutations in the Cx26 gene, responsible for DFNB1 (ref. 1), have been shown to underlie half of the cases of prelingual autosomal recessive forms of non-syndromic deafness^{2,3}. Moreover, one specific mutation (30delG, also named 35delG; ref. 3) accounts for about two-thirds of the detected mutations².

By linkage analysis of a French family (family LY1) affected by pre-lingual deafness, we previously mapped the gene responsible for a dominant form of deafness, DFNA3, to chromosome band 13q12, the same region as the gene responsible for DFNB1 (ref. 8). The affected individuals exhibited progressive hearing loss predominantly affecting the high frequencies. This led us to propose that the same defective gene could underlie both DFNB1 and DFNA3. So far, a single Cx26 mutation, resulting in a methionine-to-threonine substitution at codon 34, has been reported in a small family affected by an autosomal dominant form of deafness¹.

However, it has recently been shown that this substitution is also present in normal hearing individuals, and therefore corresponds to an asymptomatic polymorphism⁴: we have observed this polymorphism in 4 out of 190 unrelated individuals with normal hearing.

We searched for Cx26 mutations in family LY1 and in another family (LY2) from the same geographic area showing a significant log likelihood ratio score (3.6), by using polymorphic markers flanking the DFNA3 locus, namely D13S175, D13S143 and D13S115 (data not shown). Family LY1 is composed of 10 deaf individuals and 17 normal hearing individuals⁸; family LY2 is composed of 10 deaf individuals and 19 normal hearing individuals. In all the 20 affected individuals, we detected a heterozygous G-to-C transversion (Fig. 1) which creates a tryptophan-to-cysteine substitution at codon 44 (W44C). None of the 36 non-affected individuals of these two families carried the mutation. The muta-

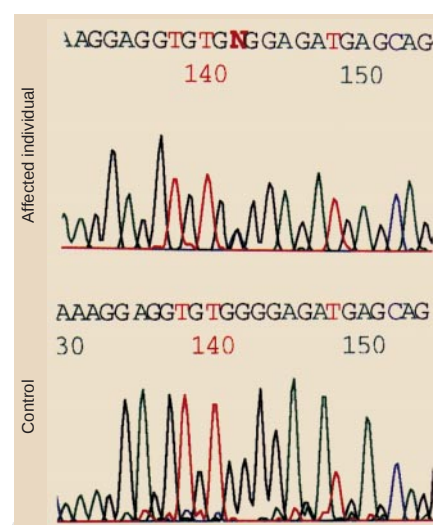


Figure 1 Sequence analysis of the Cx26 coding exon in an affected individual, showing the G-to-C transversion at codon position 44. PCR amplifications were carried out using a set of primers that allowed amplification of the entire coding sequence of Cx26 (ref. 2). Sequencing of double-stranded DNA templates was performed with fluorescent dideoxynucleotides on an Applied Biosystems (ABI373) DNA sequencer. Also, amplification products from affected individuals were subcloned and sequenced.

tion was not detected in 190 unrelated control individuals.

Connexins are gap-junction proteins; six connexin subunits assemble into a half-channel named connexon and two connexons align to make a complete intercellular channel. The topology of connexins is highly conserved, consisting of four transmembrane domains linked by one cytoplasmic and two extracellular loops, with cytoplasmic termini.

The extracellular domains of connexins have been much studied because they represent the site of interaction between connexons from adjacent cells. The tryptophan at position 44 is located in the first extracellular loop. This residue is present in 12 out of the 13 rodent connexins characterized so far; the last one harbours a tyrosine at this position, which has similar biochemical characteristics⁹. In addition, both extracellular loops contain a conserved pattern of three cysteines. These cysteines are involved in interloop disulphide bonds, and their presence has been demonstrated, by site-specific mutagenesis¹⁰, to be critical for channel function¹¹.

The presence of an additional cysteine in connexin 26 is likely to interfere with the formation of normal disulphide bonds. Therefore, the W44C mutation is expected to have a dominant-negative effect due to incorporation of mutant connexins into intercellular channels, thereby disrupting channel activity. This hypothesis is further supported by the observation that patients heterozygous for Cx26 mutations that lead to premature

protein truncation either have normal hearing or suffer only mild high-frequency hearing loss in adulthood². In another connexin gene, *Cx32*, a missense mutation at codon 44, resulting in a tryptophan-to-leucine substitution, has been detected in an individual affected by the dominant X-linked form of Charcot-Marie-Tooth disease (which causes muscular atrophy)¹².

To evaluate the extent to which DFNA3 contributes to autosomal dominant non-syndromic deafness, we searched for mutations in the coding exon of *Cx26* in individuals from 30 French families with at least three affected members. The affected individuals suffered from mild-to-profound, congenital or progressive hearing loss. No mutation in the *Cx26* coding exon was found in any of these families.

These findings establish that a mutation in *Cx26* is responsible for both a dominant and a recessive form of deafness, a situation similar to that reported for the myosin-VIIA gene which underlies DFNB2 and DFNA11 deafness^{13–15}. Although *Cx26* is a major contributor to the autosomal recessive forms of hearing loss, it seems to be responsible for only a small percentage of autosomal dominant forms.

Françoise Denoyelle

Unité de Génétique des Déficiences Sensorielles, CNRS URA 1968, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France, and Service d'ORL, Hôpital d'Enfants Armand-Trousseau, 75571 Paris cedex 12, France

Genevieve Lina-Granade, Henri Plauchu

Laboratoire de Perception et Mécanismes Auditifs, CNRS UPRESA 5020, Université Claude Bernard Lyon I, France, and Département d'ORL, Hôpital Edouard Herriot, 69437 Lyon cedex 03, France

Roberto Bruzzone

Unité de Neurovirologie et Régénération du Système Nerveux, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France

Hassan Chaïb, Fabienne Lévi-Acobas,

Dominique Weil, Christine Petit

Unité de Génétique des Déficiences Sensorielles, CNRS URA 1968, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France
e-mail: cpetit@pasteur.fr

Transgene risk is low

The prospect of a commercial release in Europe of genetically modified oilseed rape (*Brassica napus* L.) has intensified concern over the chances of transgene movement into wild relatives. We have found that potential transgene recruitment is likely to be slow and uncertain.

Brassica rapa is the most likely species occurring as natural populations to receive transgenes by introgressive hybridization¹. *B. rapa* plants growing in mixed stands with *B. napus* have yielded hybrid-seed levels of 9–93% in field studies². Similar levels (3–60% hybrid seed) were reported from *B. rapa* weeds growing in flowering oilseed rape^{3,4}. Gene flow into weedy populations depends on subsequent introgression, or dispersal into truly wild populations. This is made improbable by the likelihood that hybrid seeds will be removed during harvest and by the uncertainty of hybrids surviving farm-management practices. Introgression into *B. rapa* populations occupying their natural habitat (river and stream banks) is the most likely route for transgene spread from agriculture. To date, there are no authenticated reports of hybrids in such habitats⁴.

Gene flow will be greatest where natural populations are adjacent to large fields. *B. rapa* growing in field margins, however, may be escapes from cultivation. We studied two natural UK populations, by the River Thames, of *B. rapa* growing next to commercial oilseed rape. The first (at Hambleden Lock, Berkshire) contained 50 plants in 1996 and grew within 5 m of a 15-hectare field of *B. napus* cultivar Apex. The second (Culham, Oxfordshire) consisted of 350 plants in 1997, 1–2 m from a 12-hectare field of cultivar Apex. Seed collected from both populations was germinated and putative hybrids identified by their distinct, intermediate morphology.

The plants' hybrid status was confirmed by flow cytometry, chromosome counts and Inter-SSR (ref. 5) polymerase chain reaction (PCR) (Fig. 1). The 13,341 seeds sown from Hambleden produced 7,594 seedlings (57% germination), of which only 30 were hybrids (0.4%); 2,000 seeds sown from Culham yielded 1,053 seedlings (53% germination)

and just 16 hybrids (1.5%).

A random selection of offspring lacking the hybrid phenotype (788 plants from Hambleden, 583 from Culham) was confirmed as non-hybrid *B. rapa* by flow cytometry and Inter-SSR PCR. This sufficed to ensure that undetected hybrids did not occur at a higher frequency than detected hybrids, with a confidence of > 0.95 and > 0.99 for Hambleden and Culham, respectively⁶.

We surveyed 33.4 km of the River Thames in 1997 and identified 145 populations containing 10–1,000 *B. rapa* plants, including the Culham population and nine others (6.2% of the 145), within 360 m of oilseed rape. The density of airborne pollen beyond 360 m falls to less than 10% of that at the field margin⁷. All other populations were > 800 m from oilseed rape. We infer that there is a real, but slight, risk of gene flow in < 7% of populations in the survey area.

The Hambleden population was revisited regularly during 1997 to monitor the appearance of hybrid plants. Samples of brassica-type seedlings were removed from the site at the cotyledon/first-leaf stage and transferred to glasshouse conditions for identification. The resultant plants included 210 *B. rapa* individuals but no hybrids. Losses from the seedling population, of which we sampled less than 5% in this way, were small. The seedlings remaining at Hambleden suffered from heavy predation by slugs and snails so only 22 plants survived to maturity.

Should our results apply generally, we infer that flowering *F₁* hybrids containing neutral transgenes will appear at low frequencies in only a few natural *B. rapa* populations (0.4–1.5% hybrid seeds in < 7% of populations, with < 2% of seedlings surviving). So, although the potential for transgene recruitment is real, the process is likely to be slow and uncertain unless the transgene confers a significant selective advantage, particularly during seedling establishment.

Susan E. Scott, Mike J. Wilkinson

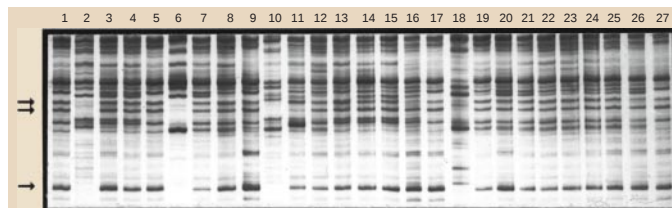
Department of Agricultural Botany, University of Reading, PO Box 221, Reading RG6 6AS, UK

1. Scheffler, J. A. & Dale, P. J. *Transgenic Res.* **3**, 263–278 (1994).
2. Jørgensen, R. B. *et al.* *Acta Hort.* **407**, 193–200 (1996).
3. Landbo, L. *et al.* *Hereditas* **125**, 89–91 (1996).
4. De Vries, F. T. *et al.* *Gorteria* (suppl.) **1**, 33 (1992).
5. Charters, Y. M. *et al.* *Theor. Appl. Genet.* **92**, 442–447 (1996).
6. Mather, K. *The Measurement of Linkage in Heredity* 26–31 (Methuen, London, 1938).
7. Timmons, A. M. *et al.* *Nature* **380**, 487 (1996).

Figure 1 Identification of

Brassica napus-*B. rapa* hybrids by Inter-SSR PCR analysis⁵. Lane 1, *B. napus* cv. Apex; lane 2, *B. rapa* plant 1; lanes 3 and 4, hybrids from

B. rapa plant 1; lane 5, *B. napus* cv. Apex; lane 6, *B. rapa* plant 5; lanes 7 and 8, hybrids from *B. rapa* plant 5; lane 9, *B. napus* cv. Apex; lane 10, *B. rapa* plant 6; lanes 11–16, hybrids from *B. rapa* plant 6; lane 17, *B. napus* cv. Apex; lane 18, *B. rapa* plant 8; lanes 19–26, hybrids from *B. rapa* plant 8; lane 27, *B. napus* cv. Apex. Arrows, *B. napus*-specific markers.



1. Kelsell, D. P. *et al.* *Nature* **387**, 80–83 (1997).
2. Denoyelle, F. *et al.* *Hum. Mol. Genet.* **6**, 2173–2177 (1997).
3. Zelante, L. *et al.* *Hum. Mol. Genet.* **6**, 1605–1609 (1997).
4. Scott, D. A. *et al.* *Nature* **391**, 32 (1998).
5. Fraser, G. R. *The Causes of Profound Deafness in Childhood* (Johns Hopkins University Press, Baltimore, 1976).
6. Petit, C. *Nature Genet.* **14**, 385–391 (1996).
7. Van Camp, G. & Smith, R. J. H. *Hereditary Hearing Loss* (Springer, New York, 1997).
8. Chaïb, H. *et al.* *Hum. Mol. Genet.* **3**, 2219–2222 (1994).
9. Bruzzone, R., White, T. W. & Goodenough D. A. *BioEssays* **18**, 709–718 (1996).
10. Dahl, G., Werner, R., Levine, E. & Rabadan-Diehl, C. *Biophys. J.* **62**, 172–182 (1992).
11. Yaeger, M. & Nicholson, B. J. *Curr. Opin. Struct. Biol.* **6**, 183–192 (1996).
12. Scherer, S. S. *et al.* in *Cell Biology and Pathology of Myelin* (eds. Juurlink, B. H. J. *et al.*) 83–102 (Plenum, New York, 1997).
13. Liu, X.-Z. *et al.* *Nature Genet.* **16**, 188–190 (1997).
14. Weil, D. *et al.* *Nature Genet.* **16**, 191–193 (1997).
15. Liu, X.-Z. *et al.* *Nature Genet.* **17**, 268–269 (1997).

Mad cows and secret agents

Fatal Protein: The Story of CJD, BSE and Other Prion Diseases

by Rosalind M. Ridley and Harry F. Baker
Oxford University Press: 1998. Pp. 249. £25, \$47.50

Adriano Aguzzi

Is it possible to condense the entire saga of the spongiform encephalopathies into a book for the lay public? When faced with reviewing this work by Rosalind Ridley and Harry Baker, two veterans of spongiform encephalopathies, I was sceptical about whether it could be done. And even after reading their book, I am still not sure of the answer.

Ridley and Baker have undertaken an audacious endeavour. In just 220 pages they recount a story that spans from the discovery of the transmissibility of sheep scrapie and the anthropological studies of Vincent Zigas and D. Carleton Gajdusek in New Guinea, through the molecular studies of the prion protein, to the political issues connected with the outbreak of bovine spongiform encephalopathy (BSE).

The last part, which describes BSE, is the best. With a public BSE inquiry under way in the United Kingdom, Ridley and Baker deliver a strong contribution by recalling the ethical dilemma of scientists and politicians as they faced a catastrophe of almost biblical proportions amid all sorts of lobbying pressures. Public opinion increasingly sensed the need to get things done quickly and, above all, to get them right.

With hindsight, it would have been all too easy to enumerate everything that went wrong in the management of the BSE crisis. Instead, Ridley and Baker provide a laudably balanced and well-informed review of what was and was not known about prions when crucial decisions were made. They emphasize that BSE was not at first thought to pose a danger, because all available evidence indicated that it was equivalent to sheep scrapie, which was rightly considered to be innocuous to humans. This part of the book may help to avert the danger of searches for a scapegoat, which is intrinsic to exercises such as this in the United Kingdom.

The book falls short of its own high standards in its attempt to link the politics with the molecules, admittedly not an easy task. Transcription, translation, protein metabolism and the molecular biology of virus replication are all mentioned and briefly explained but, apart from an extraordinarily pedestrian drawing of a cell, there are no illustrations. This is hardly of any help for those struggling to understand the biology, and will be boring for those who do understand it.



Offal waste: carcasses from cows contaminated with BSE are dyed blue and binned for disposal.

The other mildly disconcerting aspect of the book results from the obvious desire of the authors to use their opus to expose their erudition. Be it Thomas Kuhn, a whole digest of Greek philosophers, randomized quotations from their favourite movies, or platitudes such as Murphy's law, Ridley and Baker do not waste a single occasion to unload their encyclopaedic knowledge onto the poor reader. The temptation to skip these lengthy passages is almost irresistible. Although some of the digressions do not lack a sort of involuntary humour, by and large they detract from the primary objective of the book.

A very long section is devoted to familial prion diseases, and here Ridley and Baker dwell mainly on their own experience with a family carrying a pathogenic mutation of the prion gene. A very interesting bit of this section, for me at least, is the detailed exposure of the immense ethical problems of genetic

counselling in fatal diseases that manifest themselves at an adult, postreproductive age, and for which there is no treatment. The potential transmissibility of prions resident in the body of such patients adds a considerable layer of complexity.

But when it comes to the essence of the prion hypothesis, to which Ridley and Baker subscribe, they are surprisingly concise, to the point that no mention is made of the scientists who discovered the prion gene. This omission is particularly glaring because in the rest of the book intellectual property is acknowledged with an almost religious compulsiveness; even neutrinos, which arguably have little to do with spongiform encephalopathies, are not mentioned without due tribute to their discoverers.

Ridley and Baker have very strong opinions on certain aspects of prion diseases, and they make these very clear. For example, they firmly believe that scrapie is a genetic disease,

and that non-genetic transmission plays a negligible part in maintaining scrapie among sheep in the field. This is certainly a viewpoint worth discussing, but the vehemence of their argument is undoubtedly going to raise some eyebrows, especially as they repeatedly condemn the "acrimonious debate style" in the field of prion research, "which does no credit to either side".

In summary, this is an uneven book. I doubt that the detailed discussion of the maintenance of sheep scrapie in the field will interest many readers, except the authors' direct scientific competitors. Also, the lengthy accounts of their experiences with a family carrying a mutant prion gene will hardly be "enjoyed by anyone who is personally affected by prion diseases": this promise, made on the dustjacket, is rather startling considering that typical prion diseases lead to rapidly progressive dementia. But on the positive side, the first-hand chronicle of the evolution of the BSE crisis is a well-written source of information that is not easily retrievable elsewhere. □

Adriano Aguzzi is at the Institute of Neuropathology, University Hospital of Zürich, Schmelzbergstrasse 12, CH-8091 Zürich, Switzerland.
e-mail: adriano@pathol.unizh.ch

Computer evolution

Darwin Among the Machines

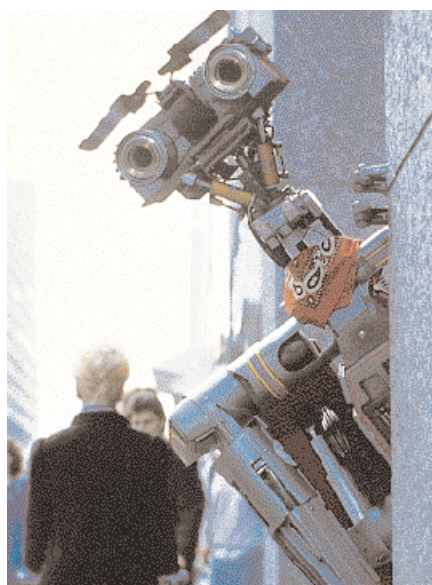
by George Dyson
Addison-Wesley/Allen Lane; 1997/1998.
Pp. 286. \$25, £20

Philip W. Anderson

George Dyson has written an extraordinarily exciting, intriguing and very idiosyncratic book. It is intensely personal in many ways, yet is so wide-ranging in subject matter and so deep in intellectual content that it is hard to describe the book in a finite compass.

In broad terms, Dyson is expressing his vision that natural selection (which he credits to Erasmus Darwin rather than to Charles) has begun to operate among our electronic machines, creating an enormous entity — the Internet — which operates in a way that no individual can either predict or control. For instance, it can strengthen useful connections and destroy unused ones on its own initiative. He sees this as beginning to embody Thomas Hobbes's dream of *Leviathan*, as he asks in the first chapter: "does this signal the emergence of intelligence or merely... a bamboo forest growing towards the light?"

But the book is more than this intriguing argument. It might be seen as at least three other books. First, there is a critical intellectual history of ideas related to evolution, artificial intelligence, and information and communication. There is also a series of



Run robot run: a robot on the loose in the film *Short Circuit II*.

anecdotal quotations by, and thumbnail sketches of, the fascinating individuals who have played a role in this history; from Pope Sylvester II and Roger Bacon in the Dark Ages, through Hobbes, Robert Hooke and William Petty in the seventeenth century, Samuel Butler, Alfred Smea and the Darwins in the nineteenth century, and Alan Turing, John von Neumann, Lewis Fry Richardson, Paul Baran and Olaf Stapledon, among many others, in more recent times.

Finally, there is also a personal narrative, involving Dyson's grandfather (a grenade specialist and musician), Dyson's own childhood days playing at the Institute for Advanced Study at Princeton, New Jersey, in the remains of von Neumann's early electronic computer, the influences of his immediate family, and his experience as a boat-builder and hermit. All of these threads are woven skilfully into a unitary whole more powerful and effective than the sum of its parts. As a work of prose and an almost perfect example of the effective literary treatment of scientific subjects, the book can hardly be faulted.

However, like his favourite seventeenth- to nineteenth-century models, Dyson has not written an unbiased scientific review or popularization in the twentieth-century sense. It is frankly an exposition of a point of view, a polemic rather than a presentation. I suspect that Dyson wants as much to spark a discussion of these topics as to present the science involved.

There is, first, an historical and institutional bias. The 'enlightened' eighteenth century gets short shrift compared with the God-centred seventeenth and the Romantic nineteenth centuries. Like Bertrand Russell, I see Hobbes's *Leviathan* rather as the precursor of statism — fascist or communist — than of the World Wide Web. Nor do I share

Dyson's enthusiasm for Samuel Butler. I felt the decades from 1960 to 1990 were slighted: Dyson focuses on the 'romantic' pre-1960 period when the computer still depended on thermionic vacuum tubes, and was in practice useful primarily for extravagant military calculations in aid of thermonuclear devastation (or lent its cachet to the unfortunate and primitive exercises in nuclear strategy of those days).

The focus on the Institute for Advanced Study is natural, but the canonization of the 'Strangelovely' RAND Corporation and the dismissive treatment of Bell Telephone Laboratories are not. (John Tukey, although credited with an "internal Bell Labs document", is described as a professor of mathematics at Princeton, a position he never held.) Dyson complains that Baran's idea of packet switching was ignored by AT&T in 1960, but the integrated circuit, which was to make it possible, had only just been conceived.

At least in part, this book may be taken as a broadside in the eternal battle of software versus hardware, of the mathematician versus the natural scientist; the mathematician is always claiming "I thought of it first" and the scientist responds with "but you didn't know what to do with it".

Dyson has a tendency to seek out the first inklings of an idea rather than its implementation. The difference between Charles and Erasmus Darwin, for instance, was that Charles had a detailed knowledge of biology and palaeontology, and so could support and clarify what were only speculations by Erasmus. Similarly, Dyson posits that the telegraph as an idea was implicit in fire signals and semaphore, but it became a usable reality only as a consequence of the detailed physical discoveries of Michael Faraday and Joseph Henry about electromagnetism. The idea of symbiogenesis became hard science rather than easy speculation as a result of modern molecular biology. My teeth were set on edge by the passage in which Dyson conflates silicon as clay, as flint, as glass fibre, and as a carrier of electrons. In general, he seems to undervalue the natural sciences in relation to mathematics.

Ultimately, when discussing Stapledon, he begins to speculate about non-corporeal intelligence using the complexity of the modern computer network. Here I was reminded of Richard Dawkins' memes: pure information systems which (since human society began) have taken advantage of more conventional means of information transfer to propagate themselves and evolve, most being pure parasites (such as fads and cults), some probably benefiting society symbiotically (such as conventional religions and social systems). Is Dyson's dream an electronic version of the meme?

The book left me intrigued and stimulated but, I must confess, not persuaded that (in Dyson's striking reference to the legend of

Bacon's Brazen Head, an automaton that was supposedly omniscient) the head has even spoken "Time Is", much less "Time Was". □

Philip W. Anderson is at the Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA.

Body of evidence

Hermaphrodites and the Medical Invention of Sex

by Alice Domurat Dreger

Harvard University Press: 1998. Pp. 268. \$35, £23.50

Roy Porter

Ignore the somewhat trendy title, which might seem to threaten yet another piece of pretentious postmodernism. This is a well-researched, sober history of a problem that Alice Dreger shows has directly affected more people than we might think and which shapes the sense of sexual identity of us all.

Knowledge that certain individuals are intersexual goes back to the dawn of history. In his *Metamorphoses*, Ovid told the story of young Hermaphroditus, the son of the gods Hermes and Aphrodite. He aroused the nymph Salmacis's passions, and she begged the gods to join them together in one body. But for a long time public attitudes to hermaphroditism involved little more than prurience, and sufferers were often reduced to displaying themselves in freak shows: Roll up! Roll up! See the bearded lady!

That had changed by the late nineteenth century. More biosexually anomalous people came to light, presumably because physicians started systematically examining children, and medical professionalism at least helped to dispel the Barnum and Bailey atmosphere. But, Dreger convincingly argues, for all their humane intentions, the doctors had agendas of their own that were not necessarily beneficial to those they hoped to help.

Doctors in nineteenth-century France and Britain were particularly struck by hermaphroditism as a problem because they had adopted rigid views on the reality of two quite separate sexes, understood as the basis of a biological reproductive division of labour. Nature meant there to be true males and pure females (and *vive la différence!*). So hermaphrodites were aberrations who needed to be (literally and metaphorically) straightened out, by being turned into the sex they were 'really' meant to be. Advances in surgery promised to help.

But 'sexing' hermaphrodites had always been, and remained, easier said than done. Intersexuals never came in any standard type. Of all the cases who turned up in Victorian doctors' surgeries, no two were alike. There were otherwise 'virile' males who had

breasts and menstruated; there were individuals with ovaries and uteruses yet luxuriant moustaches; and others seemingly had both penises and vaginas.

Dreger maintains that doctors developed a strategy for such situations: they opted to act on the basis of gonadal primacy. If someone brought up as a girl, endowed with breasts and other secondary sexual characteristics and psychologically attuned to the female role, turned out on examination to have undescended testicles, doctors would routinely identify that person as 'male'. "But, my good woman, you are a man", a physician notoriously declared in such a case.

Or, if a young male-dresser with mixed biosexual features turned out to possess anything resembling ovaries, 'he' would be deemed female and perhaps have 'his' penis removed — especially if it were not 'adequate' — and an artificial vagina constructed. The tacit assumption was the male-chauvinist one that no item of sexual anatomy could be more shameful than a non-erecting phallus.

The working rule was: 'one body, one sex'. This represented a gallant attempt by the medical men to show they could deal with the problem, and a desire to uphold socio-sexual order: after all, 'inbetweenies' could be dangerous, in schools, convents or the military. And the 'age of gonads' was reinforced by the early twentieth-century discovery of sex hormones, and the development of XX and XY chromosomal genetics.

One ironic and unforeseen consequence was that the reality of hermaphroditism



'Inbetweenie': the French hermaphrodite Marie-Madeleine Lefort, aged sixteen.



Marie-Madeleine Lefort aged sixty-five.

came to be questioned. Medical faith in the anatomical universality of two opposite sexes meant that terms such as 'pseudo' or 'spurious' hermaphroditism had to be invented to describe those poor souls to whom nature had given a motley appearance and medicine had not yet sorted out. Such labels confused more than they clarified. And, at the same time, doctors damagingly muddled intersexuality with the 'homosexuality problem' which sexologists were just demarcating. Like the 'pseudohermaphrodite', the 'homosexual' became a person in need of adjustment to his or her genital tackle; it was all a matter of anatomy rather than psychology or erotic taste.

The modern 'solution' has been to patch up 'doubtful' infants at an early age, in the understandable belief that constructing 'normality' was in the individual's best interests. Currently, however, as Dreger notes, spokespersons for intersexuals are often hostile to such 'solutions'. They resent the imposition of 'normality', with its implication of prior freakishness, and argue that sexual surgery all too often proves no better than the 'defect'. Sex is something to be negotiated not imposed.

Avoiding preachy judgementalism, Dreger shows how deeply ingrained are our assumptions about gender normality (sexual anatomy is destiny), and on how flimsy a basis they have been grounded. The book offers us all a lesson in self-awareness. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Movers and shakers

Motion Analysis of Living Cells

edited by David R. Soll and Deborah Wessels
Wiley: 1998. Pp. 298. £115, \$150

Graham A. Dunn

Cell behaviour is suddenly a hot topic. Stimulated by the mapping of the human genome, cell biologists are searching for an equally systematic approach to assigning function to gene products, and the motion analysis of living cells is a promising candidate. A large part of a cell's active genome, coding for a vast complex of signalling and regulating agents, as well as for the components of the motile machinery itself, has a potentially measurable influence on cell motility.

Until now, the stumbling block in investigating these effects has been the technical difficulty of detecting and characterizing subtle changes in cell motility, but the analysis of cell motion has now found a fairy god-mother. Computer-assisted microscopy has transformed it from a painstaking and tedious chore into an elegant method of data collection and analysis. This new volume, edited by David R. Soll and Deborah Wessels, reveals the recent upsurge of ever more sophisticated ways of characterizing cell behaviour and gives a glimpse into their often stunning visual appeal. After years of obscurity and neglect, Cinderella has finally been invited to the Cell Biology Ball.

The coverage of this multi-author work is very broad, ranging from bacterial chemotaxis and actin-based propulsion of bacterial pathogens to neural crest migration, growth-cone motility and tumour metastasis. Indeed, the topics are so disparate that it is difficult to discern a common thread. Even the theme of motion analysis indicated by the title has not been adhered to by all of the authors, although there is certainly enough of this to make the book a valuable source of methods for quantifying many aspects of cell motility.

The editors' own contributions on the motile dynamics of *Dictyostelium* amoebae reveal the amazing wealth of quantitative detail that can now be extracted from moving cells by computer-assisted microscopy. One table lists 400 numbers that were generated by the movement of a single amoeba during three minutes! This would have pleased Lord Kelvin, who once said that if one cannot reduce one's knowledge of the natural world to numbers then it is of a meagre and unsatisfactory kind. It is a shame he is no longer around to interpret them.

When the ball is over and the glittering coach turns once again into a pumpkin, Prince Charming must discover whose foot will fit into the glass slipper. Few of the techniques described have sufficient scope to qualify as general approaches, and only one paper tackles the important problem of

analysing the movement of cultured vertebrate cells. These cells are often so thinly spread in culture that their automatic recognition by computer presents a particular challenge, and the authors have developed a sophisticated, semi-automated procedure for detecting cell contours. If I were less modest, I would draw their attention to a fully automatic and intrinsically more accurate approach, based on phase-shifting interference microscopy, that has been in routine use in my own laboratory for several years.

The main argument for developing fully automated methods is that the immense diversity presented by cultured vertebrate cells, even if derived from a single clone, dictates that around a hundred or more cells must be recorded over many hours to detect a subtle experimental effect. Even semi-automated methods can soon become wearisome when dealing with the many thousands of cell outlines generated in a single experiment.

Although the automated analysis of cell behaviour is still very much in its infancy, it is clear from these pages that it is already beginning to make an impact at the core of cell research. Moreover, I suspect that it will soon make a much more profound contribution to our knowledge, not only of how cells move, but of how they communicate with each other, respond to their environment and generally cooperate to build and maintain whole organisms. More distant applications may even include routine clinical screening and diagnosis. In time, this book may well be seen as a herald of a new era in cell biology. □

Graham A. Dunn is at the MRC Muscle and Cell Motility Unit, the Randall Institute, King's College London, 26-29 Drury Lane, London WC2B 5RL, UK.

Breaking the law

Maxwell's Demon: Why Warmth Disperses and Time Passes

by Hans Christian von Baeyer
Random House: 1998. Pp. 185. \$25

Peter T. Landsberg

Adolf von Baeyer won a Nobel prize for chemistry in 1905 for his work on organic dyes and hydroaromatic compounds. His name, this book tells us, appears alongside those of other luminaries of science on a plaque outside the headquarters of the German Physical Society in the former East Berlin.

His great-grandson, Hans Christian von Baeyer, spends his summers in Paris engaged in the "vulgarisation de la science". Here he has written for the general reader a popularized introduction to thermodynamics, the second half of which gives the life history, so far, of a 'demon'. This amusing creature was introduced by James Clerk Maxwell but was named by Lord Kelvin. They thought it could help to violate the second law of thermo-

dynamics, by sorting out the fast molecules in a gas from the slow ones to produce a temperature difference in a gas that was previously at a uniform temperature.

Von Baeyer discusses the ups and downs of the demon's life, pointing out that it was thought to have been 'killed' (on paper only, of course) several times, only to emerge again with a new trick up its sleeve. Perhaps it might have been emphasized more that the second law is in fact valid in the domain of macroscopic physics, whatever manipulations one imagines to be carried out with atoms. In the course of the story, entropy and probability theory are also discussed.

This is an enjoyable book that is easy to read. Much of the history of thermodynamics related by von Baeyer is familiar, and one would not expect much novelty in this area. But the story is skilfully put together: von Baeyer has read widely, and discusses very fully Richard Feynman's ratchet-and-pawl device for possibly breaking the second law. Perhaps he misses the odd trick by not mentioning that the demon has also entered recent discussions on quantum computing. Occasionally one may question his judgement, for example when he introduces a parable in chapter 8 that does not seem to add much to the story.

The law of conservation of energy is "as close to an absolute truth as our uncertain age will permit". Well put! But readers may recall the uncertainty principle, which allows energy to be 'borrowed', breaking conservation, if this is done for a short enough time. □

Peter T. Landsberg is at the Faculty of Mathematical Studies, University of Southampton, Southampton SO17 1BJ, UK.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 10 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1996 and issued at least four separate numbers by the end of May 1998.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Peter Tallack, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

Global warming and the stability of the West Antarctic Ice Sheet

Michael Oppenheimer

Of today's great ice sheets, the West Antarctic Ice Sheet poses the most immediate threat of a large sea-level rise, owing to its potential instability. Complete release of its ice to the ocean would raise global mean sea level by four to six metres, causing major coastal flooding worldwide. Human-induced climate change may play a significant role in controlling the long-term stability of the West Antarctic Ice Sheet and in determining its contribution to sea-level change in the near future.

The West Antarctic Ice Sheet (WAIS) contains 3.8 million km³ of ice. If this entire volume was released to the ocean, the time needed to restore the ice—at the present rate of accumulation—would be more than ten thousand years. The future of WAIS has drawn attention since Mercer¹ raised the spectre of its potential disintegration under the influence of anthropogenic climate change (global warming). In an early assessment of the global warming issue, Revelle noted, “The oceans would flood all existing port facilities and other low-lying coastal structures, extensive sections of heavily farmed and densely populated river deltas of the world, major portions of the states of Florida and Louisiana, and large areas of many of the world's major cities”^{2,3}. Despite considerable progress in understanding certain aspects of WAIS in the interim, the 1995 report of the Intergovernmental Panel on Climate Change (IPCC) stated, “Our ignorance of the specific circumstances under which West Antarctica might collapse limits the ability to quantify the risk of such an event occurring, either in total or in part, in the next 100 to 1,000 years”⁴.

With countries beginning to formulate global-warming policies, the potential hazard from WAIS is too large and irreversible to be relegated to a list of imponderables for consideration at a later date. Here I review our knowledge of WAIS from the perspective of global warming in the hope of clarifying key issues and highlighting the reasons for disagreement and uncertainty. The future of WAIS will be determined by internal responses of the ice sheet to millennia-scale trends in global climate and sea level, coupled with changes in the accumulation and discharge of ice due to global warming. The probability of a major contribution to global sea-level rise is low during the twenty-first century, but increases substantially thereafter. Atmospheric accumulation of greenhouse gases over the next 100 years could irreversibly affect the future of WAIS.

Two terms that are often used without definition are: “collapse”, which I use here to mean the loss of most or all of the land-based (grounded) ice on a timescale that is much shorter than its accumulation turnover timescale⁵; and “unstable”, which means that collapse of currently grounded ice would occur following small perturbations to its boundary conditions.

Physical setting

WAIS (Fig. 1) is a marine ice sheet, with part of its ice being grounded on land below sea level, and part in the form of floating extensions called ice shelves that move seaward but are confined horizontally by the rocky coast^{6–10}. The boundary between grounded and floating ice is called the grounding line (Fig. 2). If the grounded ice were to melt, sea level would rise accordingly, by an amount proportional to the mass of grounded ice adjusted to account for the volume below sea level and for rebound of the bedrock (Fig. 2). Because ice shelves float, their melting would not displace additional water, and would not contribute to sea-level rise. Ice accumulates on WAIS through atmospheric precipitation and moves seaward from higher elevations under the influence of

gravity. Most of the drainage for WAIS is into the Ross and Weddell seas, where there are extensive ice shelves (the Ross Ice Shelf and the Filchner–Ronne Ice Shelf, respectively) and into the Amundsen Sea in the vicinity of Pine Island Bay, where extensive ice shelves are not present¹¹ (Fig. 1). If WAIS were instantaneously removed, West Antarctica would become an archipelago. Over a period of 5,000–10,000 years, the bedrock would rebound with removal of the ice load, but it would largely remain below sea level.

The stratigraphic record from the continental shelf around Antarctica indicates that WAIS began to form 15–20 million years ago^{12,13}, whereas analysis of deep-sea sediments suggests that it was about 9 million years ago¹⁴. Whether WAIS has been present continuously since that time is a matter of controversy^{14–18}. There is ample evidence of repeated advance and retreat of the ice sheet, particularly from acoustic echo studies of the ocean floor, which bear the impressions of grounded ice in motion (for reviews of this evidence, see refs 7, 19, 20). Studies of marine sediment deposited on the continental shelf indicate that WAIS was considerably larger than its present size during the Last Glacial Maximum (LGM) 13,000–24,000 years ago⁹, when the grounded ice extended at least several hundred kilometres beyond its current limit in the Ross Sea. Its ultimate limit is the edge of the continental shelf (Fig. 1) because it is nearly impossible to ground ice in the abyss²¹.

The question of whether WAIS persisted through periods that were significantly warmer than the present interglacial period remains unresolved because of uncertainties in interpretation of proxy records for sea level, local temperature, and ice extent^{9,22–24}. Based on evidence from lake sediments of a Pleistocene warm period in Antarctica, Mercer¹ surmised that WAIS may have completely disappeared at least once. Mercer also summarized evidence that global sea level stood at least 6 m higher during the last interglacial period about 120,000 years ago, but this high stand could not be tied directly to the poorly constrained chronology of the lake-sediment record. Nevertheless, these findings led to the inference that disintegration of WAIS may have caused sea level to rise at least once during a period when global mean temperature may not have reached more than 2 °C above that of today. Along with ideas, discussed below, that marine ice sheets may be inherently unstable, Mercer's findings led to concern that global warming might cause WAIS to collapse. In contrast, other major ice sheets either are largely grounded above sea level and subject to very gradual ablation from moderate warming (Greenland) or have no clear history of major, rapid, ice mass changes in the recent geological past (East Antarctica).

Ice gain and loss

The ice-sheet system gains mass through atmospheric precipitation over its entire surface and by freezing of ice on the underside of the ice shelves. It loses mass through evaporation, by melting and runoff at its surfaces (top, edges and bottom), and by calving of icebergs at the front of the ice shelves²⁵. Ice may move across the grounding line

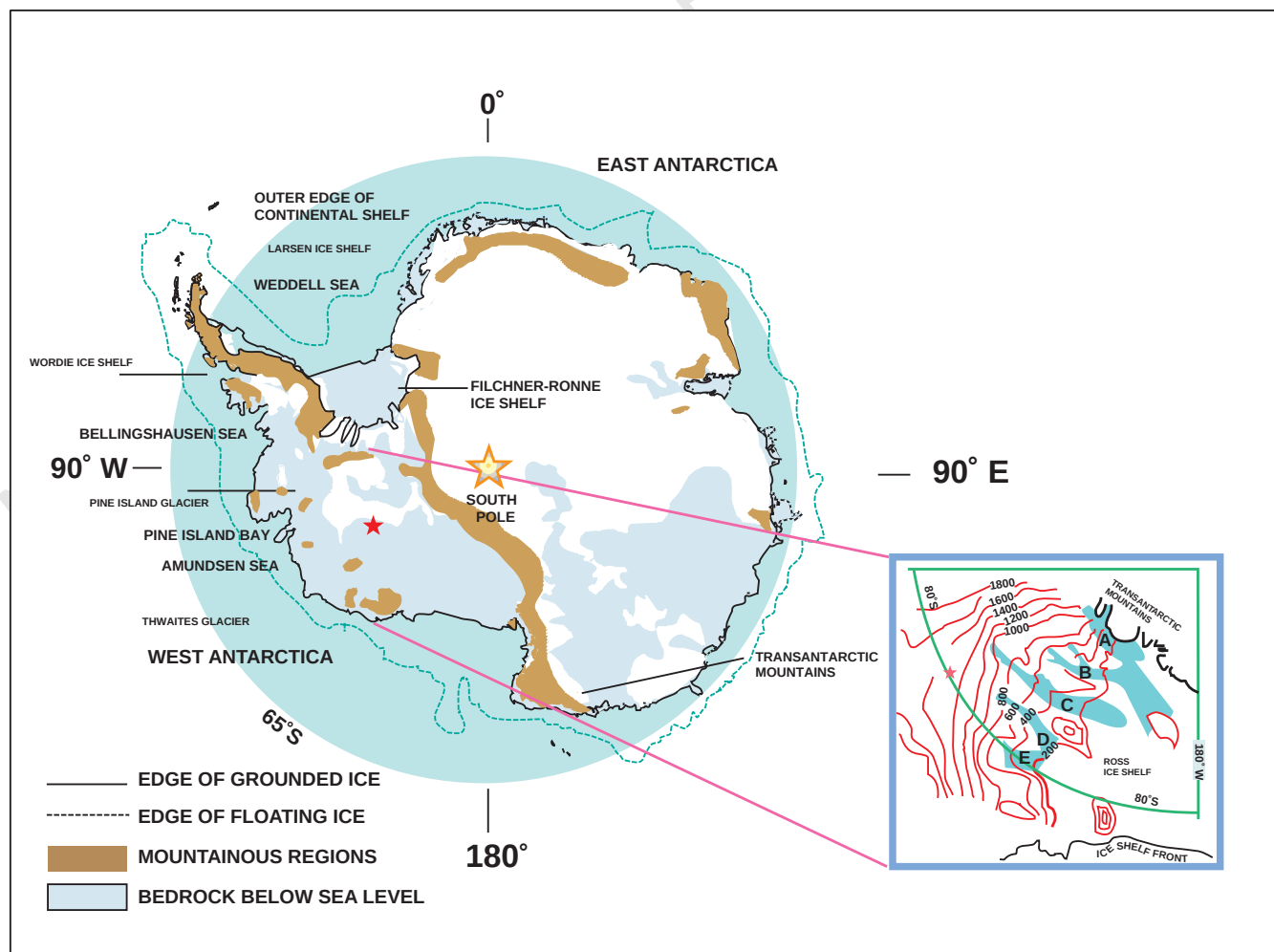
(which itself may shift in either direction), increasing or decreasing the mass of grounded ice^{3,26,27}. Mass loss from the upper surface of the grounded ice by melting is currently negligible, and loss by melting from its base is estimated to be small²⁸. Net flux of ice relative to the grounding line provides the only entry in the budget that could become large enough to account for significant net loss or gain of grounded ice, raising or lowering sea level.

The flow of ice from the grounded part of WAIS to the ocean is concentrated in fast-moving ice streams. About half of the ice flows into the Ross Ice Shelf (Fig. 1). The five ice streams in this drainage basin, each measuring 30–80 km wide and 300–500 km long, have been studied extensively using radar and photographic observation from aircraft and satellites, as well as Global Positioning and synthetic aperture radar satellite systems^{29–32}. Subglacial properties have been studied using boreholes and seismic methods^{6,33}. The ice in four of the streams is moving generally down a slope towards the grounding line at a speed averaging $\sim 0.5 \text{ km yr}^{-1}$, but with significant longitudinal (downstream) gradients. The ice streams move over sediments or bedrock and are separated from each other by ridges of relatively static ice (velocity $\sim 5 \text{ m yr}^{-1}$) of about the same thickness as the ice streams, but apparently in part frozen to their beds^{7,34–36}.

Ice streams were originally assumed to slide over bedrock, lubricated by meltwater from their base and restrained by frictional interactions at sites at the bed that are poorly lubricated. Seismic studies and evidence from boreholes and ocean-floor sediments

prove that at least two of the ice streams move over a boundary layer of unconsolidated, water-saturated sediment up to several metres thick which lubricates their motion. This layer is proposed to be deformable till that is connected to the ice above. Movement of the ice is supposed to be controlled by deformation of the till, but the degree of control is a matter of dispute^{37–42}. In an extreme view, ice streams move forward with minimal friction like a solid body floating on an inviscid fluid, in this way resembling ice shelves^{21,43}. Reality is more complex because friction also arises from ‘sticky spots’ where the till is thin or absent and bedrock pinnacles contact the ice (and where lubrication by basal meltwater may play a role)⁴⁴. Support of the ice-stream mass from the horizontal boundaries is also important⁴⁵. The relative significance of each of these sources of flow restraint is uncertain and may vary greatly, both in space and time. For example, ice stream C (Fig. 1) nearly came to a complete stop in its lower reaches about 130 years ago, possibly owing to the diversion of basal meltwater to neighbouring ice stream B, which continues to move rapidly⁴⁶.

Ice streams also drain into the Filchner–Ronne Ice Shelf in the Weddell Sea^{47–49} (Fig. 1) but these have not been as extensively characterized. A velocity of 0.4 km yr^{-1} has been measured at the grounding line of one, the Rutford Ice Stream. The Thwaites and Pine Island Glaciers are the major ice streams draining into the Amundsen Sea (Fig. 1). The latter has a small ice shelf, and velocities of $1.3\text{--}1.9 \text{ km yr}^{-1}$ have been measured inland of the grounding line.



The Thwaites Glacier appears to have a floating but unconfined ice tongue. Neither ice stream is well characterized^{11,50–53}.

The motion of the Ross and Filchner–Ronne ice shelves is restricted by contact with rises on the ocean floor and by friction with the perimeter walls of the embayments. It has been argued that these obstructions, by pinning the ice shelf, exert backpressure or buttressing which prevents collapse of a marine ice sheet that otherwise would be unstable^{27,54–57}.

Two ice shelves fringing the Antarctic Peninsula to the north of WAIS, the Wordie Ice Shelf and the northern portion of the Larsen Ice Shelf, have been disintegrating for several decades, possibly because of local atmospheric warming^{58–60}. They were not connected to major ice sheets, and their location is significantly warmer than that of the Ross and Filchner–Ronne ice shelves. The small Wordie Ice Shelf has lost about two-thirds of its area in the past thirty years, whereas the grounded ice behind it has not accelerated significantly⁶¹. But this observation may be irrelevant to the stability of a marine ice sheet like WAIS (should its ice shelves ever disintegrate) because the ice immediately behind the Wordie Ice Shelf is grounded largely at or above sea level.

Ice shelves close to the climate limit for existence ($\sim 0^{\circ}\text{C}$ in January, or -5°C mean annual isotherm) may be generally vulnerable to the rapid disintegration observed for the northern Larsen and Wordie ice shelves^{59,60,62}. The relevance of such temperature thresholds to the much larger WAIS ice shelves is a key issue. The annual mean temperature on the ice shelves of West Antarctica is -15°C to -25°C , and the mean summer value rises to about -8°C (ref. 63). A very large warming in summer, when widespread melting sometimes occurs on the ice shelves, could eventually cause them to break up and affect the stability of the grounded ice. Bentley³ argued that melting due to the direct effect of a large atmospheric warming would take thousands of years, an assertion contested by Hughes⁶⁴ and called into question by the recent behaviour of the small ice shelves. In addition, recent studies indicate that significant basal melting already occurs under the large ice shelves. Changes in ocean circulation and temperature expected to accompany global warming could decrease basal melting at some locations⁶⁵ and increase it at others²⁸, the latter circumstance providing a potential pathway for thinning and disintegration of the large ice shelves even without a summer warming of 8°C .

Is WAIS unstable?

The question of stability^{66–69} is important because the turnover time for WAIS is so long ($\sim 10^4$ yr) that a major contribution to sea-level rise over the coming centuries could not arise from a stable response

to changes in the mass balance. Hughes^{7,70} contrasted ice-stream profiles that are concave downward with early models of equilibrium ice sheets and also drew on stratigraphic studies in arguing that WAIS is inherently unstable and disintegrating as part of a regression from the LGM. Hughes also indicated mechanisms that could lead to very rapid grounding line retreat^{7,64}. Others^{3,27} have argued more convincingly that collapse could not occur within less than several hundred years.

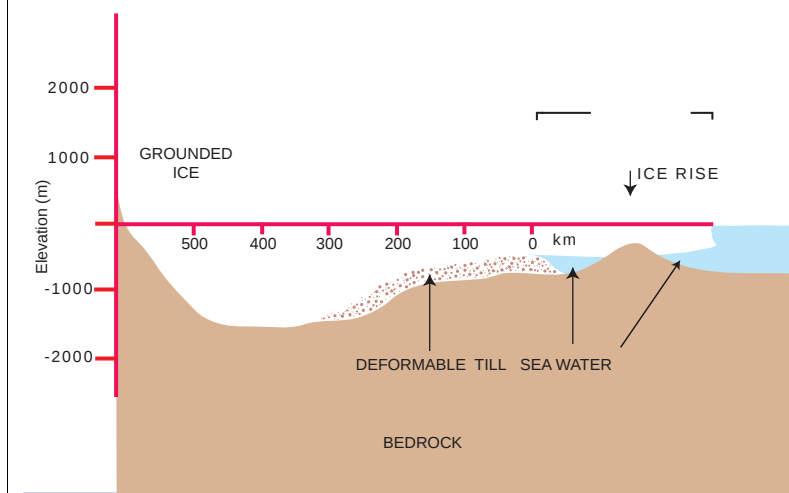
The connections among ice geometry, velocity and instability have been emphasized in many studies^{36,71,72}. But the controversy over whether the fast and variable motion of ice streams is a symptom of instability^{32,40,70} or a transient behaviour that may be damped short of collapse, or a permanent process that maintains a steady-state ice sheet by smoothly joining static inland ice with an ice shelf^{1,3,40,73–75}. This question is likely to be resolved by determining the extent of basal lubrication, for example, deformable till.

Instability in the WAIS models Weertman²⁶ provided a theoretical basis for the qualitative notion of very rapid grounding-line retreat in the absence of ice-shelf backpressure. Using a simplified model of the junction between grounded ice and a freely floating ice shelf, he demonstrated that if the bedrock slopes downwards in the inland direction away from the grounding line (a foredeepened profile, characteristic of several WAIS ice streams²⁹), then a marine ice sheet may have no stable state. If an ice shelf providing buttressing becomes unpinned owing to melting or global sea-level rise, the grounded ice would accelerate its flow, thin, and rapidly float off its bed.

This hydrostatic instability is exacerbated when it appears in WAIS models that include a simple power-law relation between the ice velocity and the ice thickness at the grounding line^{8,27}. Subsequent models use a more sophisticated (but not necessarily more realistic) description of the transition zone around the grounding line where stresses and velocities change from those typical of grounded ice to those characterizing an ice shelf^{21,76–78}. These models have found steady-state solutions even with the ice shelves arbitrarily removed. However, it is unclear whether the newer approaches eliminate the instability generally, or whether stability arises from assumptions that need further validation. None of these studies incorporate ice streams bedded on deformable till.

Models that would incorporate such ice streams with underlying lubrication, in addition to more sophisticated descriptions of the transition zone, also may eliminate the hydrostatic instability of the grounding line. But a similarly problematic junction, where basal traction changes abruptly, occurs instead at the head (inland end) of the ice stream, where relatively static ice is joined to the fast-moving ice stream^{21,79}. Describing the ice-velocity field around this junction

Figure 2 Cross-section of an ice stream and ice shelf of a marine ice sheet, indicating location of grounding line, bedrock rise on the ocean floor, and possible extent of deformable till. The thickness of the till layer, actually a few metres, is exaggerated for clarity. Sea-level rise due to collapse of WAIS was estimated by T. Hughes (personal communication) based on ref. 121.



has proved to be difficult, and simplifying assumptions can also lead to model instability (ref. 36 and discussion below), which may or may not be realistic. Furthermore, ice streams may be unstable with respect to horizontal accretion or loss of ice at their boundaries⁸⁰.

An additional problem is the wide range of timescales for the various processes that determine growth and shrinkage of an ice sheet. For WAIS, the characteristic time for restoring the ice sheet after total loss is 50,000 yr in one model⁷⁶ and for re-configuration of the bedrock in response to changes in the ice burden is 5,000–10,000 yr. The characteristic time for ice in the central part of WAIS to move to the grounding line is 1,200 yr (ref. 36) at today's velocities. Furthermore, the entire system experiences forcing by climate and sea level on all timescales. WAIS is necessarily in a continuing state of readjustment and never reaches equilibrium.

MacAyeal⁸¹ elegantly demonstrated this point by incorporating ice stream dynamics and basal till explicitly into his ice-sheet model. The till layer is heated by friction and geothermal heat from below. It is effectively insulated from atmospheric changes by the overlying ice. A thermal–gravitational instability originates when the till changes from a frozen, consolidated phase to a deformable, dilated phase in response to changes in the basal temperature as the ice thickens. Because an ice sheet that has accumulated on a frozen bed is thicker than its equilibrium dimension on a deformable bed, the phase transition causes some or all of the ice to be discharged as icebergs^{82,83}. The basal layer then refreezes and the cycle resumes with accumulation and thickening.

When this model is forced with periodic changes in climate and global sea level, deformable till develops and becomes widespread, and the ice sheet grows and discharges through ice streams on varying timescales. The magnitude and timing of the irregular and perhaps chaotic oscillations are affected by the changes in external boundary conditions, such as accumulation, that continually reset the internal clock of the ice sheet. Thus atmospheric changes influence the overall dynamics of the ice sheet by altering the transport of heat from and to the base, a process that has a very long characteristic time (of the order of 10⁴ years). One curious feature of this model is that episodes of rapid discharge are not necessarily simultaneous with interglacial periods, and collapse may occur through different WAIS drainages at widely different times. Moreover, the model reaches a steady state (after about 50,000 years) if an instantaneous warming to today's climate is imposed on an initial state consistent with LGM conditions; this effect is due to ice-shelf buttressing.

So what can the models tell us about WAIS instability? We would need a more accurate model of an ice stream and its bed and far more measurements to decide whether WAIS is now within or approaching a rapid discharge phase. Models of the entire Antarctic Ice Sheet that do not incorporate ice streams^{78,84,85} are informative but do not provide a sufficient basis for resolving the stability issue. The MacAyeal model gives important new insight into stability, but at a cost of using low resolution and limited descriptions of ice and basal layer thermodynamics and sub-ice topography. The descriptions of the ice-stream head, horizontal boundaries, and junction with the ice shelf require further development, and basal melting and freezing under the ice shelves is not taken into account. Models of individual ice streams also suffer many of these limitations.

The early idea²⁶ that WAIS is subject to hydrostatic instability and rapid grounding-line retreat seems unlikely, but this issue cannot be conclusively resolved without a better understanding of bedrock and grounding-line morphology, and ice-velocity fields, and improved models of the transition zone where the ice begins to float. The related question of how much the ice shelves influence the motion of the grounded ice remains undecided⁷⁶. Many simulations, in models with and without ice streams/deformable till, show some effect on ice velocity of removing the ice shelves completely or imposing a rapid thinning upon them^{27,43,76,86–89}. The observed seismicity beneath ice stream C (at a point 85 km upstream of the

Table 1 Empirical estimates of Antarctic Ice Sheet mass balance

	Accumulation on grounded ice	Accumulation total	Grounded ice balance	Total ice-sheet balance
Grounded ice* only	1,800 (ref. 92)		0 to +360 (ref. 92)	
	1,660 (ref. 93)		+40 to +400 (ref. 93)	
Total ice sheet†	1,528 (ref. 28)	2,144 (ref. 28)		–680 (ref. 94)

Estimates are given as Gt yr^{–1}.

* Adapted from ref. 4 (table 7.3).

† Includes accumulation on the Antarctic Peninsula and basal melting of the ice shelves.

grounding line and far outside the hypothetical transition zone) in response to tidal perturbations of the ice shelf provides direct evidence for an ice-shelf influence on the grounded ice far inland⁹⁰.

Propagation of changes in surface climate to the base of the grounded ice is very slow except near the grounding line. Thus the growth and discharge of WAIS has been partly programmed by the combination of internal dynamics and the slow response to past climate changes, for example those associated with Milankovitch cycles. In this model, WAIS may be destined to collapse over the next few centuries independent of anthropogenic climate change⁹¹.

Is today's WAIS growing or shrinking?

Until recently, empirical studies argued for a positive (growing) mass balance for both WAIS and Antarctica as a whole^{92,93}. In the absence of a validated model, this evidence provided support for the hypothesis of a stable WAIS^{3,74}. The IPCC⁴ estimated a zero Antarctic contribution to sea-level rise over the past century, and projected a small negative (about –1 cm) contribution for the twenty-first century. The latter estimate reflects a negative contribution to sea-level rise due to increased ice accumulation and an arbitrary, positive contribution (0.1–0.4 mm yr^{–1}) due to hypothetical WAIS instability. However, recent studies have inferred high melt rates under the floating ice, particularly in Pine Island Bay^{25,28,53,94}. In combination with estimates of iceberg calving and runoff, these ablation terms exceed current values for net accumulation on the entire Antarctic ice sheet by 680 Gt yr^{–1} (Table 1), although a steady state lies within the uncertainty bounds⁹⁴.

After adjustment of the mass-balance estimates above⁹³ and the new melt rates⁹⁴ to account for drainage from WAIS only (S. Jacobs, personal communication), WAIS including ice shelves has a mass balance of about –280 Gt yr^{–1}. This estimate does not exclude a steady-state situation. Lower melt rates were recently inferred for the western sector of the Ronne Ice Shelf⁹⁵. Assuming the lower values are characteristic of the entire Ronne Ice Shelf leads to a WAIS mass balance of –210 Gt yr^{–1}.

To estimate the implications of these mass-balance estimates for sea level, we assume that all the net ice-loss in the recent assessments^{28,94} is from within the grounding line. The present Antarctic contribution to global sea-level rise would then be 1.9 mm yr^{–1}, of which 0.6–0.8 mm yr^{–1} would come from WAIS. These contributions would more than account for the difference between measured sea-level rise and model calculations of changes in ocean volume over the past century^{4,25}. If unchanging, the projected increment to sea-level rise from Antarctica would be 19 cm over the coming century as compared to a total projected rise of 13–94 cm (ref. 4). The assumption that the entire negative ice balance corresponds to loss of grounded ice probably gives an upper limit, although it is conceivable that the ice shelves are gaining at the expense of the grounded ice.

Assessment of the mass balance for only grounded ice^{92,93} yields a very different perspective (Table 1). Uncertainties are high because the mass balance is derived from differences between large and uncertain accumulation/loss estimates³⁴ and because direct velocity measurements for most of the inland ice of the Weddell sector do not exist. Nevertheless, the negative total ice-sheet mass balance can be reconciled with the positive grounded mass balance estimates

only if the ice shelves are undergoing unreasonably rapid shrinkage. These issues will not be resolved until accurate ice elevations become available from satellite-based laser altimetry.

This mass-balance analysis says little about the stability of WAIS beyond undercutting the argument that it is stable because it is not losing mass. Nevertheless, WAIS-average net discharge of grounded ice would need to increase by more than a factor of six to change the framework of the global sea-level rise problem as posed by IPCC (that is, to double the best estimate of ~50 cm rise over the coming century).

How rapidly could WAIS affect sea level?

I consider here the plausible upper limit on the rate at which ice can cross the grounding line, by focusing on discharge processes and, for the moment, ignoring accumulation. Several processes were thought to limit the rate of decrease of the mass of grounded ice even in situations where hydrostatic instability was assumed to apply^{3,8,27}. Thomas *et al.*²⁷ estimated a minimum time of 400 years for disintegration of WAIS into the Ross Sea, with most of the ice loss occurring in the final century, after the ice shelf had melted as a result of a gradual warming of the ocean. Bentley³ envisaged a 'parade' of ice passing through various 'exit gates' (that is, ice stream outlets, shelf fronts, and embayment narrows) which would need to be dissipated after crossing the grounding line to prevent thickening and regrounding of the shelf. He estimated a minimum time of 500 years for collapse of WAIS through Pine Island Bay, limited by the rate at which currents clear icebergs from the embayments where the glaciers terminate.

Currents in the Ross Sea and Pine Island Bay are higher than previously thought^{28,94} and the 'parade' and 'exit gate' assumptions should also be reconsidered. Bindschadler³⁶ interprets the head of an ice stream as being an unstable junction that acts to accelerate static, upstream ice: it is assumed that the head marches backwards into the static ice rather than remaining fixed as others have implied^{40,42}. Bindschadler³⁶ also assumes that the length of an ice stream remains constant. Interpretation of the profiles of ice streams B, D and E (Fig. 1) with this model leads to a prediction that the ice streams will drain the ice back to the centre of West Antarctica and the grounding line would then retreat at about the same speed. The balance of the ice eventually would flow laterally into the ice streams, or rapidly float off the bedrock once the ice streams are fully discharged. In other words, exit gates formed by ice ridges would collapse.

There is observational support for head regression on ice streams B and C⁴⁶, but it is unclear whether the head region is actually unstable^{21,79}. Bindschadler³⁶ inferred a lower limit of 1,200 years for the remaining lifetime of WAIS, unless the apparent instability is damped. This lifetime corresponds to a mean contribution to sea-level rise of 30–50 cm per century. If this analysis were applicable to the fast-moving Thwaites and Pine Island glaciers, a collapse time of 500 years and sea-level contribution of 80–120 cm per century would be obtained.

For comparison, MacAyeal⁸¹ obtains average rates of ice release equivalent to a sea-level rise of 25 cm per century during discharge episodes, corresponding to a collapse time of 1,600–2,400 yr. As ice shelves significantly retard discharge in this model (D. R. MacAyeal, personal communication), modelled discharge rates might increase if warmer conditions than those of the present interglacial were simulated, and if basal melting were taken into account.

Achieving a high rate of discharge (collapse timescale of 500 years) would require eventual collapse of the exit gates and, in the Ross and Weddell sectors, significant acceleration of the ice streams.

Global warming and WAIS

Climate change arising from anthropogenic activities⁹⁶ may influence the timescales of WAIS accumulation and discharge through

modification of the state and behaviour of the atmosphere and ocean above and around Antarctica. So far, however, measurements are too sparse to enable the observed changes to be attributed to any such global warming.

An assessment of mostly coastal air-temperature data⁹⁷ shows a mixed pattern of modest warming and cooling (roughly $\pm 0.5^\circ\text{C}$) when data for the period 1975–1994 are compared with those from 1955–1974. The Antarctic Peninsula shows a more substantial warming of $\sim 1^\circ\text{C}$ over these periods and $\sim 2.5^\circ\text{C}$ for 1945–1990 (ref. 98). Net annual accumulation of snow has increased at several locations in the interior of East Antarctica since 1955 (ref. 99). Histories of temperature and precipitation for longer periods have been inferred from ice cores at some locations but do not provide evidence of a uniform trend^{100–102}, except near the Antarctic Peninsula¹⁰³. Some evidence of a negative trend in sea-ice extent has been presented¹⁰⁴, but the IPCC⁹⁷ assessment is that no trend has yet emerged, based on a two-decade times series beginning in 1973. However, indirect evidence of significant southward movement of the summer sea-ice edge in the two previous decades comes from examination of whaling records¹⁰⁵, and a small increase in sea-ice extent has been noted in satellite data since 1978 (ref. 106).

The effect of global warming on ice-sheet mass balance over the twenty-first century has been projected to be small and positive because it is generally assumed in ice-sheet models that the precipitation rate varies directly with the temperature-dependent saturation vapour pressure above the atmospheric inversion layer, resulting in additional accumulation^{88,107,108}. Additional accumulation is also seen in GCM experiments^{63,109}. Others have pointed out that precipitation trends also could be influenced by changes in storm patterns which cannot be projected accurately^{110,111}. For atmospheric warming that extends over millenia or is very large (for example, more than about 8°C locally), models of the entire Antarctic ice sheet show a negative mass balance due to dynamical effects and rapid melting^{84,85,88,112}. However, these models do not capture changes in ice-shelf basal melting^{28,65} that are likely to accompany changes in salinity, temperature and circulation in the circumpolar ocean arising from global warming over the next century, or the changes in ice-stream dynamics that may follow.

For example, increased precipitation would cause freshening of surface waters and reduced vertical convection and heat flux in the circumpolar ocean. Such changes would tend to decouple the surface and deep waters, causing the latter to warm. Results for the middle of the twenty-first century from the GFDL-coupled atmosphere–ocean general circulation model (AOGCM) forced with historical greenhouse gas and aerosol emissions and future emissions determined by the IPCC IS92a scenario^{96,113} (R. Stouffer, personal communication) show a spatially variable pattern of temperature change at the sea surface around Antarctica, with some areas of modest warming (0 – 1°C) and others of stronger cooling (0 – 2°C) compared to a pre-industrial control run. But the picture is quite different at depth (295 m), where warming of 0.5 – 1.5°C occurs generally and cooling is less (0 – 0.5°C) and highly localized. At 755 m, near depths of the grounding lines for several of the ice streams, warming is slightly smaller but the pattern is similar. If greenhouse-gas emissions continue to increase at a rate similar to the IS92a scenario, warming could reach 3°C at depths between a few hundred metres and one kilometre by year 2200 (ref. 114).

These results must be regarded with caution because the GFDL model has a relatively high sensitivity to doubling the atmospheric CO_2 concentration (3.7°C) and no representation of ice shelves. Projections on the space scale of one of the ice shelves, or that of the cooler, downwelling areas, are beyond the limits of credibility of model resolution. The location and persistence of downwelling areas may be critical to determining changes in basal melt rates. Another transient simulation, with the Hadley Centre AOGCM¹¹⁵ (J. F. B. Mitchell, personal communication), also projects zonal

mean warming at depth but this is smaller than the GFDL model for a given date, possibly because of the lower climate sensitivity of the former model. In these models, warming at depth goes hand-in-hand with the faster hydrological cycle that increases accumulation.

In today's climate, melting in Pine Island Bay is dominated by heat delivered by Circumpolar Deep Water (CDW), with water temperature already up to 3 °C above the *in situ* melting point. A 2-dimensional model simulation of the sub-ice circulation shows that the high melt rate could increase even further with subsurface ocean warming¹¹⁶. Based on the GCM results above, atmospheric warming and higher accumulation on Antarctica generally would be accompanied by higher basal melting at locations such as this one, where CDW dominates melting.

The Pine Island Glacier is roughly in mass balance at the present time⁵². The small ice shelf there, constrained only at its horizontal boundaries, may provide some buttressing. The Thwaites Glacier has no ice shelf to provide buttressing and its mass balance is unknown. The apparent stability of both ice streams may indicate that they are grounded on hypothetical bedrock sills^{27,29,50} (or it could be taken as evidence against hydrostatic instability). Today's high melt rate beneath Pine Island Glacier ($\sim 10 \text{ m yr}^{-1}$)⁹⁴, and the likelihood of even higher melt rates if CDW becomes warmer, raise the prospect that a sill eventually could be breached, exposing the deep subglacial basin behind it to sea water. The subsequent behaviour of these ice streams would depend on the reality of the hydrostatic instability.

Projecting the response to warming in the Ross and Weddell sectors is more complex because High Salinity Shelf Water (HSSW), formed at the sea surface freezing temperature as sea ice is produced, dominates melting today^{28,48,86,117}. AOGCM simulations give contradictory forecasts on how global warming will affect sea ice and thus HSSW. If sea-ice cover and HSSW density increase, that should speed up circulation and moderately increase basal melting. If sea ice and HSSW formation decrease, basal melting may either increase or decrease, because less dense water on the continental shelves could allow more, and warmer, CDW to penetrate the sub-ice shelf cavities¹¹⁸.

The melt rates of the Ross and Filchner–Ronne ice shelves are now estimated to be $\sim 0.22 \text{ m yr}^{-1}$ and $\sim 0.55 \text{ m yr}^{-1}$ or less, respectively. There is a range of plausible future melt rates, given the climate changes projected above, and the uncertainties in how HSSW and CDW will change^{3,86,87}. For example, a basal melt rate of 2 m yr^{-1} , far lower than those inferred currently for the Pine Island Glacier, could remove half of the Ross Ice Shelf in a little more than a century⁸⁶. Subsequently, the future of the grounded ice would depend upon the details of geology beneath the grounding line, whether hydrostatic instability is realistic, and whether the remaining shelf ice regrounds or disintegrates.

Changes in ice-shelf basal melting may also influence ice-stream dynamics in combination with thermal–gravitational instability⁸¹. Over several hundred years, surface warming cannot generally cause a phase change at the ice-stream bed. However, it is possible that the current configuration of WAIS is already unstable³⁶, or would be if the ice shelves disintegrated. Although Bentley⁷⁴ has argued that the frequency of unstable configurations is low (~ 0.001) if they occur randomly, episodes of rapid discharge appear to be correlated⁸¹. Regression from the LGM occurred relatively recently, and perhaps rapidly^{7,9}.

If we assume that WAIS is indeed in or near an unstable configuration, and that the world will continue to warm, then how may the rate of discharge change? First, thinning or removing the ice shelves may cause an increase of ice-stream velocities that would speed discharge by a factor of 2–3 (ref. 43). The higher velocities on the Pine Island Glacier and the Thwaites Glacier could be interpreted as supporting this proposition, although explanations apart from an absence of buttressing, such as higher basal shear stress, are at least as plausible. In addition, this velocity

increase would enhance frictional heating and extend the domain of lubricated basal conditions. We cannot assess the potential significance of the second process. But the first could presumably increase discharge into the Ross Sea by a factor of 2–3 (because absent hydrostatic instability, the rate of grounding-line retreat may be comparable to the ice-stream velocity), and shrink the collapse time to 400–600 years once discharge increases (based on ref. 36).

Future scenarios and policy implications

It is not possible to place high confidence in any specific prediction about the future of WAIS. Nevertheless, policy makers are confronted by three scenarios that span the range of plausible outcomes for WAIS, assuming continued growth in greenhouse-gas emissions according to rates characteristic of most of the IPCC 'IS92' emissions projections⁹⁶. The following ice-sheet scenarios reflect the assumption that there is a gradual dynamic response to removal of the ice shelf, no dynamic response, and very rapid dynamic response, respectively.

(1) Ice streams remain active. Basal melt rates gradually increase, thinning and eventually eliminating the Ross Ice Shelf in 200 years. As the shelf thins, ice-stream velocities increase, leading to higher discharge rates. As a result, the Antarctic contribution to sea-level rise remains positive (0–19 cm per century), despite ice accumulation increases due to atmospheric warming. Eventually, static ice drains into the ice streams or floats. The duration of the collapse process is 500–700 years and the WAIS contribution to sea-level rise reaches ~ 60 –120 cm per century. Collapse might have occurred anyway, but more slowly, if the ice shelves had remained in place.

(2) Ice streams slow as a result of internal ice sheet readjustments. Discharge of grounded ice decreases even if the ice shelves thin. Velocities of the Thwaites and Pine Island glaciers do not change much because they are already moving fast. The Antarctic contribution to sea-level rise turns increasingly negative^{88,119,120} until warming of about 8 °C has occurred, far more than expected over the next century.

(3) The Ross Ice Shelf thins and eventually disintegrates, and grounding lines recede rapidly owing to hydrostatic instability. The WAIS contribution to sea-level rise increases gradually over two centuries as the shelf thins and ice-stream velocities increase. Then sea level rises sharply as the bulk of WAIS is discharged in a matter of 50–200 years, perhaps limited by the speed at which embayments are cleared of floating ice. The whole process takes 250–400 years. WAIS also could discharge in the vicinity of Pine Island Bay by hydrostatic instability if one of the ice streams there is barely stable today: for example, resting on a sill. Given the clearing time of the embayment and inferred currents^{3,94}, the timescale would be about 200 years after melting destabilized the grounding line.

I assess the relative likelihood of scenario (1) to be highest of the three (but with low confidence), even though it would require considerable acceleration of ice discharge, because it best reflects the current activity of the ice streams and the probable mass balance of WAIS. Scenarios intermediate between (1) and (2) are conceivable, with lower sea-level rise than for (1): for example, if partial collapse is followed by readjustment. Progress on understanding WAIS over the past two decades has enabled us to lower the relative likelihood of scenario (3) but not to eliminate it entirely. One outcome that may be put aside for the moment, because no convincing model of it has been presented, is a sudden collapse that causes a 4–6 m sea-level rise within the coming century.

Scenario (1) would result in total rates of sea-level rise beyond year 2100 that are double or triple the median values projected by the IPCC for the twenty-first century. Global average rates of rise would exceed the recent historical value by about an order of magnitude and be comparable to today's effective rates in subsiding deltaic areas such as Bangladesh. Such rates of change pose a serious

challenge to societal adaptive capability and could prove devastating to coastal ecosystems.

Scenarios (1) and (3) would ultimately result in 4–6-m sea-level rise, accompanied by all the disruptive connotations noted by Revelle. Almost all sea-level rise would occur beyond the twenty-first century. However, these outcomes are predicated on changes in basal melt rates that could accompany global warming of only a few degrees, warming that could be determined by emissions that occur during the twenty-first century⁹⁶. Given the long residence time of atmospheric carbon dioxide, and thermal inertia in the oceans, decisions being considered now could irreversibly affect WAIS in the distant future. For example, if global emissions follow any of several of the IPCC 'IS92' scenarios⁹⁶ over the twenty-first century, sufficient climate forcing may be created that, no matter what the policy responses are afterwards, substantial thinning of the Ross Ice Shelf during the twenty-second century would be assured. □

Michael Oppenheimer is at the Environmental Defense Fund, 257 Park Avenue South, New York, New York 10010, USA.

1. Mercer, J. H. Antarctic ice and Sangamon Sea level. *Int. Assoc. Sci. Hydrol. Symp.* **79**, 217–225 (1968).
2. Revelle, R. R. in *Changing Climate* 441–448 (National Academy Press, Washington DC, 1983).
3. Bentley, C. R. The West Antarctic ice sheet: diagnosis and prognosis. *Proc. Carbon Dioxide Research Conf. USA* (NTIS, Springfield, VA, 1982).
4. Intergovernmental Panel on Climate Change *1995: The Science of Climate Change* (eds Houghton, J. T. *et al.*) 359–405 (Cambridge Univ. Press, 1996).
5. Rodhe, H. in *Global Biogeochemical Cycles* (eds Butcher, S. S., Charlson, R. G., Orians, G. H. & Wolfe, G. V.) 55–72 (Academic, London, 1992).
6. Bentley, C. R. in *Geology of Antarctica* (ed. Tinguely, R. J.) 335–364 (Oxford Univ. Press, 1991).
7. Hughes, T. Is the West Antarctic ice sheet disintegrating? *J. Geophys. Res.* **78**, 7884–7910 (1973).
8. Thomas, R. H. & Bentley, C. R. A model for Holocene retreat of the West Antarctic ice sheet. *Quat. Res.* **10**, 150–170 (1978).
9. Denton, G. H. *et al.* Late Wisconsin and early Holocene glacial history, inner Ross embayment, Antarctica. *Quat. Res.* **31**, 151–182 (1989).
10. Thomas, R. H. Ice shelves: a review. *J. Geophys. Res.* **24**, 273–286 (1979).
11. Kellogg, T. B. & Kellogg, D. E. Recent glacial history and rapid ice stream retreat in the Amundsen sea. *J. Geophys. Res.* **92**, 8859–8864 (1987).
12. Shipp, J. *et al.* High- to intermediate-resolution seismic stratigraphic analysis of mid-late-Miocene to Pleistocene strata in eastern Ross sea: implications for changing glacial/climatic regime. *Terra Antarctica* **1**, 381–384 (1994).
13. Anderson, J. B. & Bartek, L. R. Cenozoic glacial history of the Ross sea revealed by intermediate resolution seismic reflection data combined with drill site information. *Antarctic Res. Ser.* **56**, 231–263 (1992).
14. Kennett, J. P. & Barker, P. F. *Proc. Ocean Drilling Prog.* **113**, 937–960 (1990).
15. Kennett, J. P. & Hodell, D. A. Evidence for relative climatic stability of Antarctica during the early Pliocene: a marine perspective. *Geografiska Annaler* **75**, A, 205–220 (1993).
16. Warnke, D. A., Bonnie, C. A., Hodell, D. A. Major deglaciation of East Antarctica during the late Pliocene? Not likely from a marine perspective. *Mar. Micropaleon.* **27**, 237–251 (1996).
17. Scherer, R. P. Quaternary and tertiary microfossils from beneath ice stream B: evidence for a dynamic West Antarctic ice sheet history. *Paleogeogr. Paleoclimatol. Paleocool.* **90**, 395–412 (1991).
18. Burrows, L. J. Is there direct evidence for late quaternary collapse of the West Antarctic ice sheet? *J. Glaciol.* **39**, 491–494 (1993).
19. Stuiver, M. *et al.* in *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T. J.) 319–426 (Wiley-Interscience, New York, 1981).
20. Anderson, J. B. *et al.* Evidence for a grounded ice sheet on the Ross sea continental shelf during the late Pleistocene and preliminary paleodrainage reconstruction. *Antarctic Res. Ser.* **57**, 39–62 (1992).
21. Hindmarsh, R. C. A. in *Ice in the Climate System* (ed. Peltier, W. R.) **1**, 67–99 (Springer, Berlin, 1993).
22. Lambeck, K. & Nakada, M. Constraints on the age and duration of the last interglacial period and on sea level variations. *Nature* **357**, 125–128 (1992).
23. Howard, W. R. A warm future in the past. *Nature* **388**, 418–419 (1997).
24. Neumann, A. C. & Hearty, P. J. Rapid sea-level changes at the close of the last interglacial (substage 5e) recorded in Bahamian island geology. *Geology* **24**, 775–778 (1996).
25. Jacobs, S. S. Is the Antarctic ice sheet growing? *Nature* **360**, 29–33 (1992).
26. Weertman, J. Stability of the junction of an ice sheet and an ice shelf. *J. Glaciol.* **13**, 3–11 (1974).
27. Thomas, R. H., Sanderson, T. J. O. & Rose, K. E. Effect of climatic warming on the West Antarctic ice sheet. *Nature* **277**, 355–358 (1979).
28. Jacobs, S. S. *et al.* Melting of ice shelves and the mass balance of Antarctica. *J. Glaciol.* **38**, 375–387 (1992).
29. Bentley, C. R. Antarctic ice streams: a review. *J. Geophys. Res.* **92**, 8843–8858 (1987).
30. Shabtaie, S., Whillans, I. M. & Bentley, C. R. The morphology of ice streams A, B, and C, West Antarctica, and their environs. *J. Geophys. Res.* **92**, 8865–8883 (1987).
31. Bindshadler, R. A. & Scambos, T. A. Satellite-image-derived velocity field of an Antarctic ice stream. *Science* **252**, 242–246 (1991).
32. Bindshadler, R. & Vornberger, P. Changes in the West Antarctic ice sheet since 1963 from declassified satellite photography. *Science* **279**, 689–692 (1998).
33. Englehardt, H. *et al.* Physical conditions at the base of a fast moving Antarctic ice stream. *Science* **248**, 57–59 (1990).
34. Shabtaie, S. & Bentley, C. R. West Antarctic ice streams draining into the Ross ice shelf: configuration and mass balance. *J. Geophys. Res.* **92**, 1311–1336 (1987).
35. Whillans, I. M., Bolzan, J. & Shabtaie, S. Velocity of ice streams B and C, Antarctica. *J. Geophys. Res.* **92**, 8895–8902 (1987).
36. Bindshadler, R. Actively surging West Antarctic ice streams and their response characteristics. *Ann. Glaciol.* **24**, 409–414 (1997).

37. Blankenship, D. D. *et al.* Seismic measurements reveal a saturated porous layer beneath an active Antarctic ice stream. *Nature* **322**, 54–57 (1986).
38. Alley, R. B. *et al.* Deformation of till beneath ice stream B, West Antarctica. *Nature* **322**, 57–59 (1986).
39. Blankenship, D. D. *et al.* Till beneath ice stream B. 1. Properties derived from seismic travel times. *J. Geophys. Res.* **92**, 8903–8911 (1987).
40. Alley, R. B. *et al.* Till beneath ice stream B. 4. A coupled ice-till flow model. *J. Geophys. Res.* **92**, 8931–8940 (1987).
41. Kamb, B. & Englehardt, H. in *Proc. of the International Symposium held at St. Petersburg, September 1990* (IAHS Pub. No. 208, 145–154, 1991).
42. Anandakrishnan, S. *et al.* Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations. *Nature* (in the press).
43. MacAyeal, D. R. Large-scale ice flow over a viscous basal sediment: theory and application to ice stream B, Antarctica. *J. Geophys. Res.* **94**, 4071–4087 (1989).
44. Alley, R. B. In search of ice-stream sticky spots. *J. Glaciol.* **39**, 447–454 (1993).
45. Echelmeyer, K. A. *et al.* The role of the margins in the dynamics of an active ice stream. *J. Glaciol.* **40**, 527–538 (1994).
46. Anandakrishnan, S. & Alley, R. B. Stagnation of ice stream C, West Antarctica by water piracy. *Geophys. Res. Lett.* **24**, 265–268 (1997).
47. Doake, C. S. M. *et al.* Glaciological studies on Rutford ice stream, Antarctica. *J. Geophys. Res.* **92**, 8951–8960 (1987).
48. Jenkins, A. & Doake, C. S. M. Ice-ocean interaction on Ronne ice shelf, Antarctica. *J. Geophys. Res.* **96**, 791–813 (1991).
49. Smith, A. M. Basal conditions on Rutford ice stream West Antarctica, from seismic observations. *J. Geophys. Res.* **102**, 543–552 (1997).
50. Crabtree, R. D. & Doake, C. S. M. Pine island glacier and its drainage basin: results from radio-echo sounding. *Ann. Glaciol.* **3**, 65–70 (1982).
51. Lucchitta, B. K. *et al.* Antarctic glacier-tongue velocities from Landsat images: first results. *Ann. Glaciol.* **17**, 356–366 (1993).
52. Lucchitta, B. K., Rosanova, C. E. & Mullins, K. F. Velocities of Pine island glacier, West Antarctica, from ERS-1 SAR images. *Ann. Glaciol.* **21**, 277–283 (1995).
53. Jenkins, A. *et al.* Glaciological and oceanographic evidence of high melt rates beneath Pine island glacier, West Antarctica. *J. Glaciol.* **43**, 114–121 (1997).
54. Thomas, R. H. The dynamics of marine ice sheets. *J. Glaciol.* **24**, 167–177 (1979).
55. Thomas, R. H. The creep of ice shelves: interpretation of observed behaviour. *J. Glaciol.* **12**, 55–70 (1973).
56. Mercer, J. H. West Antarctic ice sheet and CO₂ greenhouse effect: a threat of disaster. *Nature* **271**, 321–325 (1978).
57. Lingle, C. S. A numerical model of interactions between a polar ice stream and the ocean: application to ice stream E, West Antarctica. *J. Geophys. Res.* **89**, 3523–3549 (1984).
58. Doake, C. S. M. & Vaughan, D. G. Rapid disintegration of the Wordie ice shelf in response to atmospheric warming. *Nature* **350**, 328–330 (1991).
59. Rott, H., Skvarca, P. & Nagler, T. Rapid collapse of northern Larsen ice shelf, Antarctica. *Science* **271**, 788–792 (1996).
60. Vaughan, D. G. & Doake, C. S. M. Recent atmospheric warming and retreat of ice shelves on the Antarctic peninsula. *Nature* **379**, 328–331 (1996).
61. Vaughan, D. G. Implications of the break-up of Wordie ice shelf, Antarctica, for sea level. *Antarctic Sci.* **5**, 403–408 (1993).
62. Doake, C. S. M. *et al.* Breakup and conditions for stability of the northern Larsen Ice Shelf, Antarctica. *Nature* **391**, 778–780 (1998).
63. Ohmura, A., Wyld, M. & Bengtsson, L. A possible change in mass balance of Greenland and Antarctic ice sheets in the coming century. *J. Clim.* **9**, 2129–2135 (1996).
64. Hughes, T. The stability of the West Antarctic ice sheet: what has happened and what will happen. *Proc. Carbon Dioxide Res. Conf. USA* 1982 (NTIS, Springfield, VA, 1982).
65. Nicholls, K. W. Predicted reduction in basal melt rates of an Antarctic ice shelf in a warmer climate. *Nature* **388**, 480–482 (1997).
66. Weertman, J. The stability of ice-age ice sheets. *J. Geophys. Res.* **66**, 3783–3792 (1961).
67. Nye, J. F. The response of glaciers and ice-sheets to seasonal and climatic changes. *Proc. R. Soc. Lond. A* **256**, 559–584 (1960).
68. Hollin, J. T. On the glacial history of Antarctica. *J. Glaciol.* **4**, 173–195 (1962).
69. Robin, G. deQ. & Adie, R. J. in *Antarctic Research* (eds Priestley, E., Adie, R. J. & Robin, G. deQ.) 100–117 (Butterworths, London, 1964).
70. Hughes, T. West Antarctic ice streams. *Rev. Geophys. Space Sci.* **15**, 1–46 (1977).
71. Shabtaie, S., Bindshadler, R. A. & MacAyeal, D. R. Mass-balance studies of ice streams A, B, and C, West Antarctica, and possible surging behavior of ice stream B. *Ann. Glaciol.* **11**, 137–149 (1988).
72. Lingle, C. S. & Brown, T. J. in *Dynamics of the West Antarctic Ice Sheet* (eds van der Veen, C. J. & Oerlemans, J.) 249–285 (Reidel, Dordrecht, The Netherlands, 1987).
73. Weertman, J. Glaciology's grand unsolved problem. *Nature* **260**, 284–286 (1976).
74. Bentley, C. R. Rapid sea-level rise soon from West Antarctic ice sheet collapse? *Science* **275**, 1077–1078 (1997).
75. Alley, R. B. & Whillans, I. M. Changes in the West Antarctic ice sheet. *Science* **254**, 959–963 (1991).
76. van der Veen, C. J. Response of a marine ice sheet to changes at the grounding line. *Quat. Res.* **24**, 257–267 (1985).
77. Herterich, K. in *Dynamics of the West Antarctic Ice Sheet* (eds van der Veen, C. J. & Oerlemans, J.) 185–202 (Reidel, Dordrecht, The Netherlands, 1987).
78. Huybrechts, P. The Antarctic ice sheet during the last glacial-interglacial cycle: a three-dimensional experiment. *Ann. Glaciol.* **14**, 115–119 (1990).
79. Barillon, V. & MacAyeal, D. R. Steady flow of a viscous ice stream across a no-slip/free-slip transition at the bed. *J. Glaciol.* **39**, 167–185 (1993).
80. Van der Veen, C. J. & Whillans, I. M. Model experiments on the evolution and stability of ice streams. *Ann. Glaciol.* **23**, 129–137 (1996).
81. MacAyeal, D. R. Irregular oscillations of the West Antarctic ice sheet. *Nature* **359**, 29–32 (1992).
82. MacAyeal, D. R. Binge/purge oscillations of the Laurentide ice sheet as a cause of the north Atlantic's Heinrich events. *J. Paleoclimatol.* **8**, 775–784 (1993).
83. Greve, R. & MacAyeal, D. R. Dynamic/thermodynamic simulations of Laurentide ice-sheet instability. *Ann. Glaciol.* **3**, 328–335 (1996).
84. Verbitsky, M. Y. & Oglesby, R. J. The CO₂-induced thickening/thinning of the Greenland and Antarctic ice sheets as simulated by a GCM (CCM1) and an ice-sheet model. *Clim. Dynam.* **11**, 247–253 (1995).
85. Fastook, J. L. & Prentice, M. A finite-element model of Antarctica: sensitivity test for meteorological mass-balance relationship. *J. Glaciol.* **40**, 167–175 (1994).

86. Lingle, C. S. in *Glaciers, Ice Sheets and Sea Level: Effect of a CO₂-Induced Climate Change* 317–330 (US Dept of Energy, Washington DC, 1985).
87. Thomas, R. H. in *Glaciers, Ice Sheets and Sea Level: Effect of a CO₂-Induced Climate Change* 301–316 (US Dept of Energy, Washington DC, 1985).
88. Huybrechts, P. & Oerlemans, J. Response of the Antarctic ice sheet to future greenhouse warming. *Clim. Dynam.* **5**, 93–102 (1990).
89. Budd, W. F., Jessen, D., Mavrakakis, E. & Coutts, B. Modelling the Antarctic ice-sheet changes through time. *Ann. Glaciol.* **20**, 291–297 (1994).
90. Anandakrishnan, S. & Alley, R. B. Tidal forcing of basin seismicity of ice stream C, West Antarctica, observed far inland. *J. Geophys. Res.* **102**, 15183–15196 (1997).
91. Alley, R. B. & MacAyeal, D. R. West Antarctic ice sheet collapse: chimera or clear danger? *Antarc. J.-Rev.* **1993** 59–60 (1993).
92. Budd, W. F. & Smith, I. N. in *Glaciers, Ice Sheets and Sea Level: Effect of a CO₂-Induced Climate Change* 172–177 (US Dept of Energy, Washington DC, 1985).
93. Bentley, C. R. & Giovinetto, M. B. in *International Conference on the Role of the Polar Regions in Global Change: Proceedings of a Conference Held June 11–15, 1990 at University of Alaska, Fairbanks* (eds Weller, G., Wilson, C. L. & Serverin, B. A. B.) 481–488 (Univ. of Alaska, Fairbanks, 1991).
94. Jacobs, S. S., Hellmer, H. H. & Jenkins, A. Antarctic ice sheet melting in the southeast Pacific. *Geophys. Res. Lett.* **23**, 957–960 (1996).
95. Nichols, K. W., Makinson, K. & Johnson, M. R. New oceanographic data from beneath Ronne ice shelf, Antarctica. *Geophys. Res. Lett.* **24**, 167–170 (1997).
96. Intergovernmental Panel on Climate Change *1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 9–49 (Cambridge Univ. Press, 1996).
97. Intergovernmental Panel on Climate Change *1995: The Science of Climate Change* (eds Houghton, J. T. et al.) 133–192 (Cambridge Univ. Press, 1996).
98. King, J. C. Recent climate variability in the vicinity of the Antarctic peninsula. *Int. J. Climatol.* **14**, 357–369 (1994).
99. Mosley-Thompson, E. et al. Recent increase in South Pole snow accumulation. *Ann. Glaciol.* **21**, 131–138 (1995).
100. Jouzel, J. Climatic information over the last century deduced from a detailed isotopic record in the South Pole snow. *J. Geophys. Res.* **88**, 2693–2703 (1993).
101. Mosley-Thompson, E. in *Climate since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 572–591 (Routledge, New York, 1992).
102. Isaksson, E. et al. A century of accumulation and temperature changes in Dronning Maud Land, Antarctica. *J. Geophys. Res.* **101**, 7085–7094 (1992).
103. Peel, D. A. in *Climate since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 549–571 (Routledge, New York, 1992).
104. Jacka, T. H. & Budd, W. F. in *International Conference on the Role of the Polar Regions in Global Change: Proceedings of a Conference held June 11–15, 1990 at University of Alaska, Fairbanks* (eds Weller, G., Wilson, C. L. & Serverin, B. A. B.) 63–70 (Univ. of Alaska, Fairbanks, 1991).
105. de la Mare, W. K. Abrupt mid-twentieth-century decline in Antarctic sea-ice extent from whaling records. *Nature* **389**, 57–59 (1997).
106. Cavalieri, D. J. et al. Observed hemispheric asymmetry in global sea ice changes. *Science* **278**, 1104–1106 (1997).
107. Robin, G. deQ. Ice cores and climate change. *Phil. Trans. R. Soc. Lond. B* **280**, 143–168 (1977).
108. Lorius, C. et al. A 150,000-year climatic record from Antarctic ice. *Nature* **316**, 591–595 (1985).
109. Thompson, S. L. & Pollard, D. Greenland and Antarctic mass balances for present and doubled atmospheric CO₂ from the GENESIS version-2 global climate model. *J. Am. Meteorol. Soc.* **10**, 871–900 (1997).
110. Kapsner, W. R. et al. Dominant influence of atmospheric circulation on snow accumulation in Greenland over the past 18,000 years. *Nature* **373**, 52–54 (1995).
111. Bromwich, D. Ice sheets and sea level. *Nature* **373**, 18–19 (1995).
112. Huybrechts, P. Formation and disintegration of the Antarctic ice sheet. *Ann. Glaciol.* **20**, 336–340 (1994).
113. Haywood, J. M. et al. Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations. *Geophys. Res. Lett.* **24**, 1335–1338 (1997).
114. Manabe, S. & Stouffer, R. J. Multiple-century response of a coupled ocean-atmosphere model to an increase of atmospheric carbon dioxide. *J. Climate* **7**, 5–23 (1994).
115. Murphy, J. M. & Mitchell, J. F. B. Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part II: spatial and temporal structure of response. *J. Clim.* **8**, 57–80 (1995).
116. Hellmer, H. H., Jacobs, S. S. & Jenkins, A. Oceanic erosion of a floating Antarctic glacier in the Amundsen sea. *Antarctic Res. Ser.* **75**, (in the press).
117. MacAyeal, D. R. Thermohaline circulation below the Ross ice shelf: a consequence of tidally induced vertical mixing and basal melting. *J. Geophys. Res.* **89**, 597–606 (1984).
118. Grosfeld, K., Gerdes, R. & Determann, J. Thermohaline circulation and interaction between ice shelf cavities and the adjacent open ocean. *J. Geophys. Res.* **102**, 15595–15610 (1997).
119. Budd, W. F. & Simmonds, I. in *International Conference on the Role of the Polar Regions in Global Change: Proceedings of a Conference held June 11–15, 1990 at University of Alaska, Fairbanks* (eds Weller, G., Wilson, C. L. & Serverin, B. A. B.) 489–494 (Univ. of Alaska, Fairbanks, 1991).
120. Gregory, J. M. & Oerlemans, J. Sea-level rise from glacier melt over the next century. *Nature* **391**, 474–476 (1998).
121. *Antarctica: Glaciological and Geophysical Folio* (ed. Drewry, D. J.) sheets 3 and 4 (Scott Polar Research Institute, Cambridge, UK, 1983).
122. Bindshadler, R., Vornberger, P. L. & Shabtaie, S. The detailed net mass balance of the ice plain on Ice Stream B, Antarctica: a geographic information system approach. *J. Glaciol.* **39**, 471–482 (1993).

Acknowledgements. I thank R. Alley, R. Bindshadler, C. Bentley, R. Hindmarsh, P. Huybrechts, S. Jacobs, A. Jenkins, C. Raymond and I. Whillans for their comments on drafts of this manuscript. R. J. Stouffer and J. F. B. Mitchell kindly provided data in advance of publication.

KNOW YOUR COPYRIGHT RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

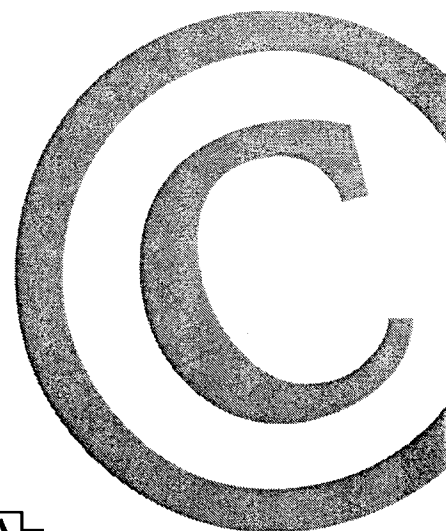
If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP

Telephone: 0171 436 5931
Fax: 0171 436 3986



RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor

Linda M. McLatchie^{*†}, Neil J. Fraser^{*}, Martin J. Main^{*}, Alan Wise^{*}, Jason Brown^{*}, Nicola Thompson^{*}, Roberto Solari^{*}, Melanie G. Lee^{*} & Steven M. Foord^{*}

^{*} Receptor Systems and Cell Biology Units, GlaxoWellcome Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

Calcitonin-gene-related peptide (CGRP) and adrenomedullin are related peptides with distinct pharmacological profiles. Here we show that a receptor with seven transmembrane domains, the calcitonin-receptor-like receptor (CRLR), can function as either a CGRP receptor or an adrenomedullin receptor, depending on which members of a new family of single-transmembrane-domain proteins, which we have called receptor-activity-modifying proteins or RAMPs, are expressed. RAMPs are required to transport CRLR to the plasma membrane. RAMP1 presents the receptor at the cell surface as a mature glycoprotein and a CGRP receptor. RAMP2-transported receptors are core-glycosylated and are adrenomedullin receptors.

The calcitonin family of peptides comprises five known members: calcitonin, amylin, two calcitonin-gene-related peptides (CGRP1 and CGRP2) and adrenomedullin (ADM) (Fig. 1). Calcitonin is involved in the control of bone metabolism and is also active in the central nervous system (CNS). Amylin also has specific binding sites in the CNS and is thought to regulate gastric emptying and have a role in carbohydrate metabolism. The CGRP peptides differ from each other by three amino acids and have so far proved to be indistinguishable in their biological activities. These activities include the regulation of neuromuscular junctions, of antigen presentation within the immune system, of vascular tone and of sensory neurotransmission^{1,2}. ADM, the most recently discovered member, is also a potent vasodilator^{1,2}. ADM has specific receptors on astrocytes³ and its messenger RNA is upregulated in CNS tissues that are subject to ischaemia⁴.

The calcitonin family peptides probably act through seven-transmembrane-domain, G-protein-coupled receptors (GPCRs). The gene encoding the calcitonin receptor has been cloned⁵. The calcitonin receptor is homologous to GPCRs in 'family B', which typically recognize regulatory peptides such as secretin, glucagon and vasoactive intestinal polypeptide (VIP). Later, CRLR was identified^{6,7}. It has 55% overall identity with the calcitonin receptor. Other members of the calcitonin peptide family were candidate ligands for CRLR but CRLR seemed not to be a CGRP receptor⁸. Two related members of the 'family A' class of GPCRs, RDC1 and G10D, were then identified as receptors for CGRP and ADM, respectively^{9,10}. This was unexpected, as although RDC1 and G10D are homologous to each other, they are very different from the calcitonin receptor; conversely, CGRP, ADM and calcitonin are related. Later, a cytosolic protein, RCF (receptor component factor), from the guinea-pig organ of Corti was expression-cloned because of its ability to induce CGRP-mediated responses in *Xenopus laevis* oocytes¹¹. Finally, a single HEK293 cell line engineered to express CRLR also bound to CGRP and showed CGRP-mediated responses¹².

Isolation and cloning of RAMP1

We chose to clone the gene encoding the human CGRP receptor by an expression-cloning strategy using a SK-N-MC cell complementary

DNA library. SK-N-MC cells are derived from a human neuroblastoma, bind ¹²⁵I-labelled CGRP1 and respond to CGRP by increasing levels of intracellular cyclic AMP¹³. The cDNA library was transcribed *in vitro* and pools of complementary RNA were injected into *Xenopus* oocytes with cRNA encoding the cystic fibrosis transmembrane regulator (CFTR). CFTR contains a cAMP-activated chloride channel that can be used as a sensitive read-out for receptors that couple positively to adenylyl cyclase¹⁴. Human CGRP1 was used throughout and is referred to as CGRP.

Xenopus oocytes have an endogenous CGRP receptor¹⁵, detected through the CFTR as a small inward current after CGRP application (23 ± 15 nA at -60 mV, $n = 15$). Similar, small CGRP responses were recorded in most oocytes expressing the pools of SK-N-MC cRNA. However, from a single pool CGRP responses were recorded that were larger than the endogenous response. This pool of clones was subdivided repeatedly according to their CGRP responses and a single cDNA was isolated that encoded a 148-amino-acid protein which we have called receptor-activity modifying protein 1 (RAMP1). In oocytes expressing RAMP1, large, dose-dependent responses to CGRP (1330 ± 315 nA, $n = 15$) were recorded (Fig. 2, inset). The responses were inhibited by the CGRP receptor antagonist CGRP₈₋₃₇, with an estimated pA_2 of 8.4, typical of a type 1 CGRP receptor¹ (Fig. 2). RAMP1-expressing oocytes showed no significant responses to ADM, calcitonin or amylin. *Xenopus* oocytes contain endogenous adenosine, VIP and β -adrenergic receptors, each of which will couple to the CFTR, but expression of RAMP1 had no effect on the responses of these receptors when they were activated (data not shown). In this system, RAMP1 appears to be specific for CGRP-mediated responses.

The hydrophobicity plot of the RAMP1 protein indicates that it has an amino-terminal signal sequence and single putative transmembrane domain close to the carboxy terminus (see below). This

ADM	YRQSMNNFQGLRSFGCRFGTCTVQKLAHQIYQFTDKDKDVA ¹ PRSKI ¹ SPQGY
CGRP1	ACDTATCVTHRLAGLLSRSGGVVKNFVPTNVGSKAF
CGRP2	ACNTATCVTHRLAGLLSRSGGMVKSNFVPTNVGSKAF
AMYLIN	KCNTATCATQRLANFLVHSSNNFQAILSS ¹ TNVGSNTY
CALCITONIN	CGNLSTCMLGTYYTQD-----FNKPHFT ¹ FPQTAIGVGAP

Figure 1 The amino-acid sequences of the calcitonin family peptides, aligned for comparison.

[†] Present address: The Novartis Institute for Medical Sciences, 5 Gower Place, London WC1E 6BN, UK.

was unexpected as we thought that the CGRP receptor would be a seven-transmembrane-domain GPCR. However, RAMP1 is not, by itself, a CGRP receptor. The effector concentration for half-maximal response (EC_{50}) of 10 nM for CGRP in RAMP1-expressing oocytes is at least tenfold lower than that for the CGRP receptor in SK-N-MC cells. Expression of RAMP1 in mammalian cells (human embryonic kidney (HEK)293T, COS7 or Swiss3T3 cells) did not induce cAMP responses to CGRP or specific binding to 125 I-labelled CGRP1. We speculated that RAMP1 induced CGRP-mediated responses in oocytes because it could potentiate the endogenous CGRP receptor. RAMP1 might thus be expected to potentiate a mammalian CGRP receptor in the same way.

CRLR requires RAMP1 for functional expression

The GPCRs RDC1 and CRLR have been reported to be receptors for CGRP. We expressed both GPCRs in *Xenopus* oocytes, but neither altered the endogenous response to CGRP (Fig. 3a). In the same experiments, oocytes expressing the calcitonin receptor responded well to calcitonin (>500 nA; data not shown). When CRLR was expressed with RAMP1, a larger CGRP response was recorded than if RAMP1 had been expressed alone ($3,111 \pm 806$ nA versus $1,330 \pm 315$ nA; Fig. 3). In contrast, expression of RDC1 with RAMP1 did not alter the oocytes' responses to CGRP (data not shown).

We studied the interaction between CRLR and RAMP1 in HEK293T cells (a subclone of HEK293 cells that has been engineered to express the SV40 large T antigen). HEK293T cells do not express endogenous CGRP or calcitonin receptors (unlike some HEK293 lines)¹⁶. Neither RAMP1 nor CRLR induced significant responses to CGRP when transfected alone, but expression of both produced cells that responded to CGRP by increasing intracellular cAMP levels and binding specifically to 125 I-labelled CGRP1 (Fig. 3b, c). RAMP1 and CRLR reconstituted a CGRP receptor that displayed virtually identical pharmacology to the CGRP receptor in SK-N-MC cells (Table 1) and in a CRLR-expressing cell line¹². RDC1 (ref. 10) did not induce binding or responses to CGRP in HEK293T cells, with or without co-expression of RAMP1 (data not shown).

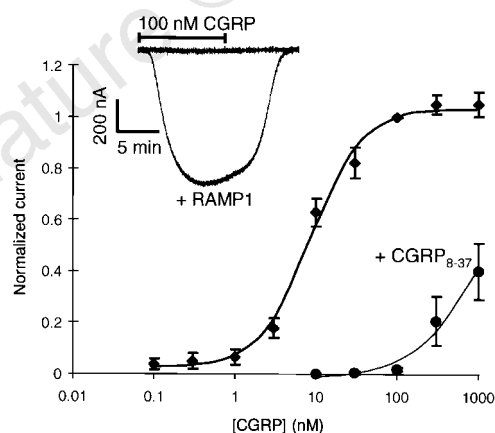


Figure 2 *Xenopus* oocytes expressing RAMP1 show large and dose-dependent responses to CGRP. Oocytes were injected with CFTR cRNA alone or in combination with RAMP1 cRNA. Recordings were made 4 days post-injection. Inset shows responses produced by 100 nM CGRP (bar) at a holding potential of -60 mV (inward current shown by downward deflection). Main graph shows dose-response curves for peak inward current under control conditions and in the presence of 100 nM CGRP₈₋₃₇ (all in the presence of RAMP1 and CFTR). The data from each oocyte were normalized according to their responses to 100 nM CGRP. From curves fitted to these data, we calculated EC_{50} values as 9 ± 1 nM for CGRP (control) and $1,400 \pm 400$ nM (with 100 nM CGRP₈₋₃₇). RAMP1 was selective for CGRP over human amylin, calcitonin and ADM, $n = 4-7$. All data are mean $\pm$ s.e.m.

The requirement for CRLR and RAMP1 to reconstitute a CGRP receptor explains why it has been difficult to expression-clone, its failure to function in oocytes, and the observation that CRLR can only function as a CGRP receptor in certain cell lines^{8,12,16} (which presumably express endogenous RAMP1; see below).

Mechanism of RAMP1 activity at CRLR

We proposed that CRLR represents the CGRP receptor and that RAMP1 transports it to the cell surface. To test this, we analysed HEK293T cells transiently transfected with Myc-epitope-tagged CRLR by fluorescence-activated cell sorting (FACS). Four per cent of the cells transfected with Myc-CRLR showed significant cell-

Table 1 Characterization of CGRP and ADM receptors in recombinant and native cell lines

Peptide	125 I-CGRP1 binding* IC ₅₀ (nM)		125 I-rat ADM binding† IC ₅₀ (nM)	
	SK-N-MC	CRLR-RAMP1	NG108-15	CRLR-RAMP2
CGRP1	0.1	0.1	1,966	1,300
CGRP2	0.01	0.01	130	300
ADM ₁₋₆₂	0.4	9	0.21	0.75
ADM ₁₃₋₆₂	3	100	25.7	17
ADM ₂₂₋₆₂	600	400	60	15
CGRP ₈₋₃₇	0.8	0.3	120	75
Amylin	NSE	NSE	NSE	NSE
Calcitonin	NSE	NSE	NSE	NSE

* Binding of 125 I-labelled CGRP1 was determined in membranes prepared from SK-N-MC cells and from Swiss 3T3 cells constructed to express CRLR and RAMP1. Typically, the specific binding of 125 I-labelled CGRP1 was 3,500 and 1,500 c.p.m. for SK-N-MC and Swiss 3T3 membranes, respectively, using 50 μ g of membrane protein per point.

† Binding of 125 I-labelled rat ADM was determined in membranes prepared from NG108-15 cells and from HEK293T cells transiently transfected with cDNA encoding CRLR and RAMP2. 20 μ g membrane protein was used from NG108-15 cells and 10 μ g protein from HEK293T cells. Typically, the specific binding of 125 I-labelled rat ADM was 2,000 and 4,000 c.p.m. for NG108-15 and HEK293T membranes, respectively. Binding data were analysed by nonlinear regression using the computer program Alfit. NSE, no significant effect. All determinations were made three times and s.e.m. are less than 10% of the mean in each case.

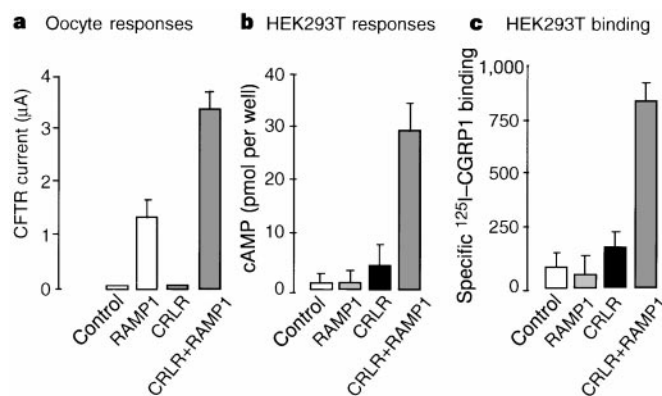


Figure 3 Expression of RAMP1 with CRLR in oocytes and HEK293T cells. **a**, In oocytes, CGRP elicits a small response through an endogenous receptor (control) that is markedly potentiated on expression of RAMP1. CRLR has no effect when expressed alone but allows very large responses when expressed with RAMP1. **b**, In mammalian HEK293T cells there is no endogenous CGRP receptor. RAMP1 alone has no significant ability to produce a cAMP response to CGRP, and neither does CRLR when transfected alone. Expression of RAMP1 with CRLR in HEK293T cells conferred cAMP responses to CGRP. **c**, Specific binding of 125 I-labelled CGRP1 to plasma membranes from HEK293T cells transfected with the same cDNAs and at the same time as those shown in **b**.

surface immunoreactivity. This increased to 22% if RAMP1 was expressed with Myc-CRLR (Fig. 4a, top). In the reverse experiment, HEK293T cells were transiently transfected with Myc-epitope-tagged RAMP1, Myc-RAMP1 could not be detected on the cells unless they were permeabilized. However, when CRLR was expressed with Myc-RAMP1, more than 40% of the cells expressed Myc-RAMP1 on the surface (Fig. 4b, bottom). Following permeabilization, 70% of cells showed immunoreactivity towards Myc-RAMP1 or Myc-CRLR, indicating high transfection efficiency. Thus, expression of RAMP1 and CRLR leads to both proteins being presented at the cell surface.

We performed crosslinking studies with the chemical bis[sulpho-succinimidyl]suberate (BS^3) with the aim of irreversibly linking ^{125}I -labelled CGRP1 to its receptor on the surface of HEK293T cells transfected with RAMP1 and CRLR. We studied whole cells before solubilization and analysis on SDS-polyacrylamide gel electrophoresis (SDS-PAGE). As expected, ^{125}I -labelled-CGRP1 could be cross-linked only to the surface of cells transfected with both proteins. After SDS-PAGE, ^{125}I -labelled CGRP1 was crosslinked to two proteins with relative molecular masses (M_r) of 66K and 17K (Fig. 4c). Both associations could be prevented by incubation with 1,000 nM unlabelled CGRP. An M_r of 66K is consistent with that of the native human CGRP receptor^{17,18}. The smaller, 17K band¹⁷ could be RAMP1. The crosslinking of ^{125}I -labelled CGRP1 to both proteins indicates a close association, as BS^3 can crosslink primary amine groups that are ~ 20 Å away from each other.

HEK293T cells expressing Myc-CRLR produced a distinct band of 58K after SDS-PAGE and immunoblotting (Fig. 5a, c) and no 66K band. Myc-RAMP1 ran as a monomer of M_r 14K. At higher expression levels, multimers of Myc-RAMP1 could be detected even under reducing conditions. Expressing Myc-CRLR with increasing amounts of Myc-RAMP1 caused the 58K band to disappear and a diffuse band of about 66K to appear (Fig. 5a). The 66K band size was consistent with that of the protein crosslinked to ^{125}I -labelled

CGRP1 on the surface of cells transfected with RAMP1 and CRLR (Fig. 4c). The appearance of the 66K band correlated with specific binding of ^{125}I -labelled CGRP1 to cell membranes (Fig. 5b). We thought that RAMP1 might be required for the terminal glycosylation of CRLR.

Endoglycosidases F and H can be used to differentiate between N-linked glycoproteins. When the 58K band (representing CRLR expressed in the absence of RAMP1) was treated with endoglycosidase F, a single immunoreactive species of M_r 48K was seen after SDS-PAGE. Thus, the 58K form is a glycoprotein (Fig. 5c, lane 2). The 66K form (derived from expression of CRLR with RAMP1) was also reduced to a 48K form by treatment with endoglycosidase F, showing that the additional mass units gained when RAMP1 is expressed with CRLR represent carbohydrate residues (Fig. 5c, lane 3). The 66K form is resistant to endoglycosidase H, indicating that CRLR has been terminally glycosylated, an event normally associated with transit through the Golgi complex and the production of mature glycoproteins (Fig. 5c, lane 5). The 58K form is sensitive to endoglycosidase H, indicating that it has not been terminally glycosylated. The mechanism of RAMP1 activity is consistent

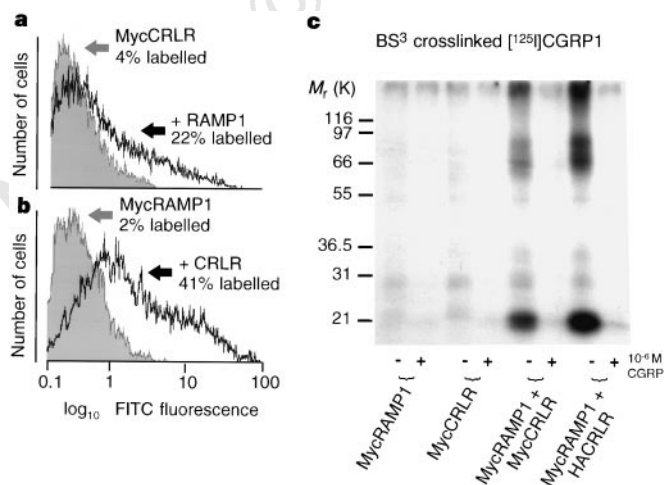


Figure 4 RAMP1 and CRLR are localized together on the cell surface when they are co-expressed. **a**, FACS analysis of HEK293T cells transiently expressing Myc-epitope-tagged CRLR ($1 \mu\text{g ml}^{-1}$ cDNA) with or without RAMP1 ($5 \mu\text{g ml}^{-1}$ cDNA). The number of cells sorted as showing cell-surface Myc-CRLR increased from 4% to 22% of total when RAMP1 was co-transfected. **b**, FACS analysis of Myc-epitope-tagged RAMP1 transfected cells with or without CRLR. The number of cells sorted as showing cell-surface Myc-RAMP1 increased from 2% to 41% of total when CRLR was co-transfected. **c**, Transfection of CRLR with RAMP1 enabled BS^3 specifically to crosslink ^{125}I -labelled CGRP1 to proteins of 66K and 17K on the surface of intact HEK293T cells. CRLR and RAMP1 alone had no effect. Similar results were obtained with HA-tagged CRLR.

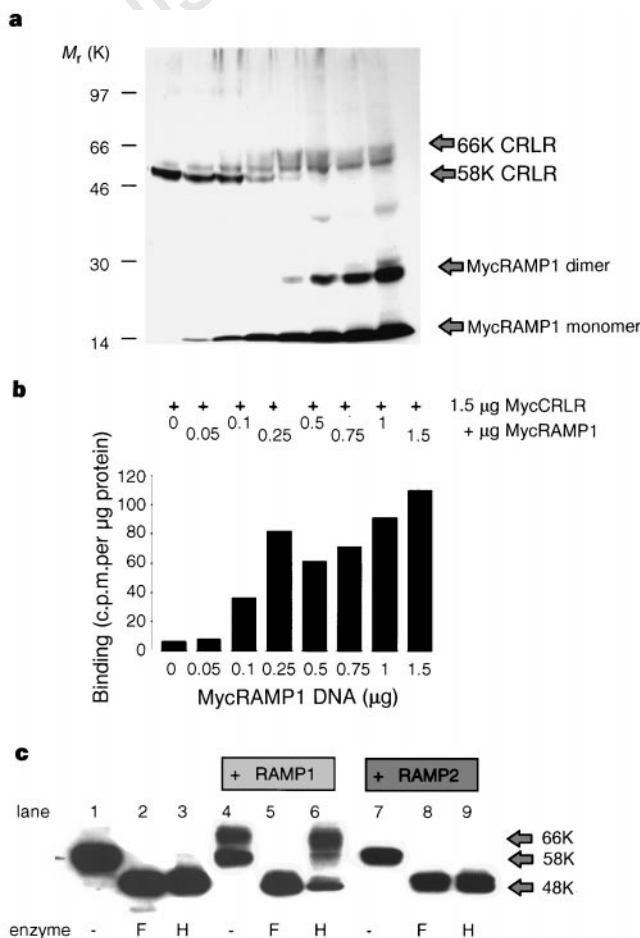


Figure 5 RAMP1 acts to increase the relative molecular mass of CRLR at the same time as specific binding of ^{125}I -labelled CGRP1 becomes apparent. Cells were collected 48 h after transfection with $1.5 \mu\text{g}$ Myc-CRLR or with Myc-CRLR and increasing amounts of Myc-RAMP1. P2 particulate membrane fractions were prepared for reducing SDS-PAGE, followed by immunoblotting with anti-Myc antiserum (**a**) or by radioligand binding studies with ^{125}I -labelled CGRP1 (**b**). **c**, The 58K CRLR species (produced without RAMP1) is a glycoprotein that is reduced by either endoglycosidase F or endoglycosidase H to a 48K band. The 66K species produced after co-expression with RAMP1 reduced to a 48K band after endoglycosidase F treatment but was resistant to endoglycosidase H, indicating that it is a mature glycoprotein.

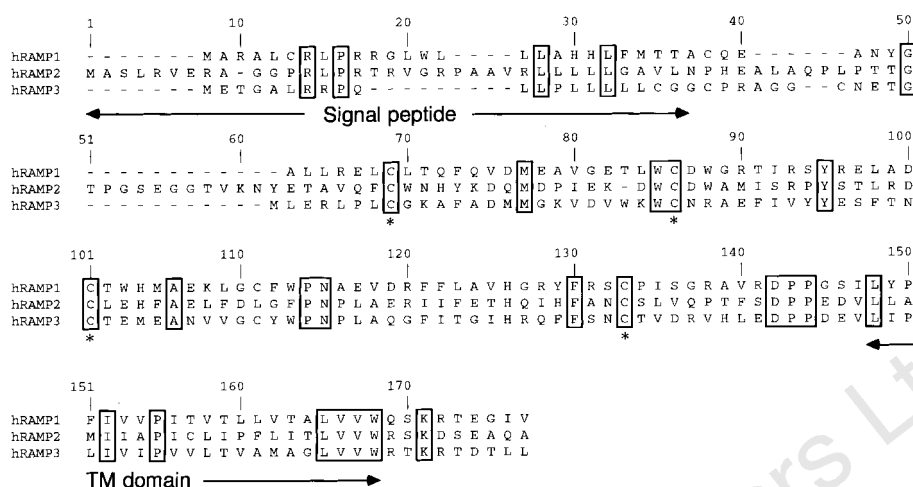


Figure 6 Amino-acid sequences of human RAMPs 1, 2 and 3, aligned for comparison. Putative signal sequences and transmembrane (TM) domains are

indicated. Conserved amino acids are boxed and conserved cysteine residues are indicated by asterisks.

with a transport event: CRLR functions as a CGRP receptor after it has been presented at the plasma membrane as a mature glycoprotein. However, during this process RAMP1 is also expressed at the plasma membrane and it may contribute more directly to the function of the CGRP receptor.

The RAMP family of proteins

The human RAMP1 sequence was partially represented by a number of expressed sequence tags (ESTs) in public databases. Searches of these databases also led to the discovery of the partial sequences of two further RAMP1-like proteins, which we have named RAMP2 and RAMP3. A full-length RAMP2 cDNA was isolated from the SK-N-MC cell cDNA library by colony hybridization using a probe amplified by polymerase chain reaction (PCR). A full-length RAMP3 cDNA was isolated by conventional and rapid amplification of cDNA ends (RACE) PCR from human spleen mRNA. Figure 6 shows the amino-acid sequences of human RAMP1, RAMP2 and RAMP3, aligned for comparison. The RAMPs are about 31% identical and about 56% similar to each other. We have identified rodent homologues of RAMPs, but have been unable to find invertebrate homologues.

The hydrophobicity plots of the RAMP amino-acid sequences are very similar, despite the low sequence homologies. In the N termini, four cysteine residues are conserved, suggesting some common secondary structure. RAMPs are type I transmembrane proteins, with an extracellular N terminus and a cytoplasmic C terminus, an orientation confirmed by FACS experiments (Fig. 4b). Both RAMP1 and RAMP3 have a pair of basic residues close to their short, cytoplasmically orientated C terminus. This pair of basic residues is similar to an endoplasmic-reticulum-retrieval motif¹⁹. However, RAMP2 has a basic-acidic pair of residues rather than a dibasic pair at this position. The functional importance of these residues will be tested.

RAMP mRNA cell and tissue expression

We prepared northern blots using mRNA extracted from SK-N-MC, Kelly (another human cell line derived from a neuroblastoma) and HEK293T cells. Of these cell lines, only SK-N-MC cells exhibit specific binding to ¹²⁵I-labelled CGRP1 or cAMP responses to CGRP¹³ (see Fig. 3 for HEK293T cells; data not shown for Kelly cells). A CRLR-specific probe revealed the expected transcripts¹² of 7.5, 5.5 and 3.5 kilobases (kb) in SK-N-MC cells, but no transcripts in Kelly or HEK293T cells (data not shown). The blots were also hybridized with probes specific for RAMPs 1, 2 and 3. A single

transcript of 0.8 kb was seen for RAMP1 and RAMP2 in each cell line; however, RAMP3 mRNA was detected in Kelly cells only (Fig. 7a). Quantification of the RAMP1 and RAMP2 transcripts showed that SK-N-MC and Kelly cells express higher levels of RAMPs 1 and 2 than do HEK293T cells, and that the ratio of RAMP1 to RAMP2 was greatest in SK-N-MC cells.

Figure 7b shows multiple human tissue northern blots examined with the same probes. The transcript sizes for RAMP mRNAs are the same as seen in the cell lines. The RAMP1 gene is expressed in many tissues, including the uterus, bladder, brain, pancreas and gastrointestinal tract. RAMPs 2 and 3 have similar, but not identical, tissue distributions and are expressed strongly in the lung, breast, immune system and fetal tissues.

The data from the cell lines show that RAMPs are not necessarily expressed with CRLR. The distributions of RAMP1 and CRLR

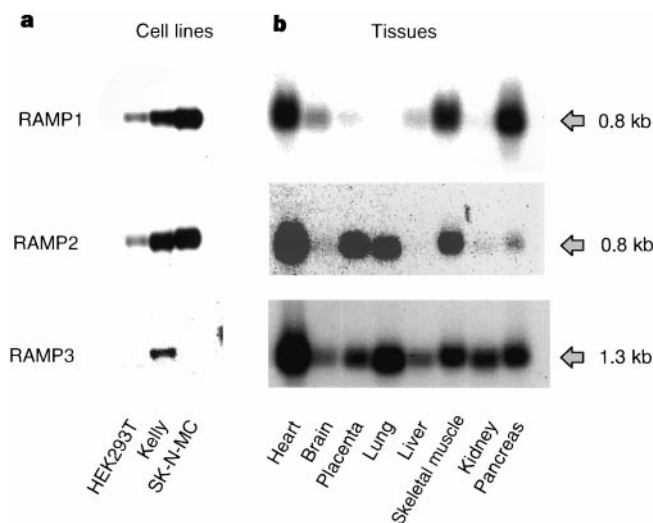


Figure 7 A northern blot analysis of mRNA encoding RAMPs 1, 2 and 3 in cell lines and tissues. **a**, Northern blots of mRNA extracted from human HEK293T, Kelly and SK-N-MC cell lines and hybridized to probes specific for RAMPs 1, 2 and 3. Of these cell lines, only SK-N-MC cells showed either responses to CGRP or CRLR transcripts. mRNA loading was equivalent for each cell line, as judged by comparable expression of glyceraldehyde 3-phosphate dehydrogenase (not shown). **b**, Multiple tissue northern blots representing human tissues, probed as described above.

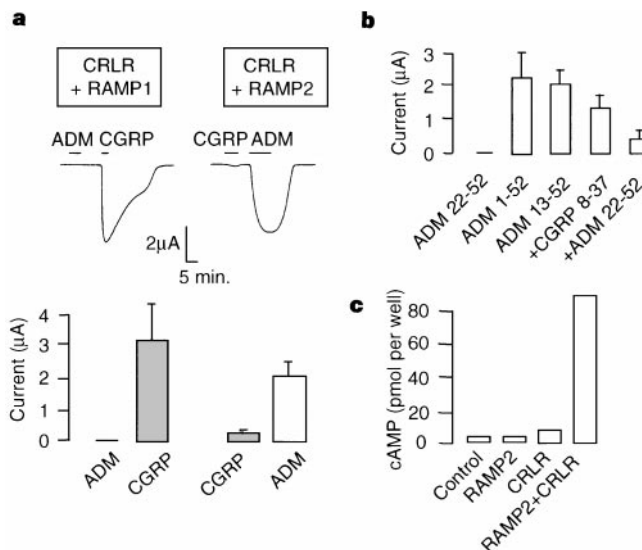


Figure 8 Expression of RAMP2 plus CRLR in oocytes and in HEK293T cells. **a**, Top, representative traces showing responses to 10 nM ADM₁₃₋₅₂ and 100 nM CGRP in oocytes expressing RAMP1 (left panel) and RAMP2 (right panel) plus CRLR. Bottom, cumulative data from at least eight oocytes for each peptide. Expression of CRLR with RAMP1 in oocytes produced robust responses to 100 nM CGRP but no significant response to 10 nM ADM₁₃₋₅₂. In contrast, oocytes from the same batches expressing CRLR with RAMP2 showed robust responses

to 10 nM ADM₁₃₋₅₂ and small responses to 100 nM CGRP. **b**, A pharmacological analysis of the ADM receptor produced in oocytes by expression of CRLR with RAMP2. ADM₁₋₅₂ and ADM₁₃₋₅₂ are effective agonists whereas ADM₂₂₋₅₂ and CGRP₈₋₃₇ are not agonists and significantly antagonized the response to 10 nM ADM₁₃₋₅₂ at 100 nM. *n* was 4–8 oocytes for each treatment. **c**, RAMP2 and CRLR expressed together in HEK293T cells allow new cAMP responses to 10 nM ADM₁₃₋₅₂. RAMP2 and CRLR had no effect when expressed alone.

overlap in many tissues and *in situ* localization of the mRNA encoding the two proteins should define which cells express CGRP receptors. CGRP receptors are found in many tissues and always at relatively low levels. High levels of mRNA encoding CRLR^{6,12,16} are found in the lung; however, the levels of CGRP receptors in the lung are unremarkable. We detected little RAMP1 mRNA in the lung; in contrast, RAMP2 and RAMP3 mRNAs were particularly abundant. We therefore examined interactions between CRLR and the other RAMP proteins.

RAMP2 and CRLR generate an ADM receptor

In contrast to RAMP1, RAMPs 2 and 3 were unable to potentiate the oocyte response to CGRP. Similar results were obtained in HEK293T cells, in which neither RAMP2 nor RAMP3 allowed the CRLR to function as a CGRP receptor. However, we demonstrated by FACS analysis that RAMPs 2 and 3 were as effective at transporting epitope-tagged CRLR to the cell surface as was RAMP1 (data not shown). SDS-PAGE analysis of epitope-tagged CRLR expressed with either RAMP2 (Fig. 5c) or RAMP3 (not shown) showed mainly 58K proteins and not the 66K form associated with binding to ¹²⁵I-labelled CGRP1 and responses to CGRP. This 58K species was reduced to a 48K species by treatment with endoglycosidase F or H. Thus, RAMP2 facilitates the transport of CRLR to the plasma membrane, but CRLR transported is a 58K, endoglycosidase-H-sensitive glycoprotein.

As RAMPs 2 and 3 enabled a form of CRLR that is not a conventional CGRP receptor to be presented at the cell surface, we proposed that the receptor might be activated by another ligand. We studied ADM for several reasons: first, ADM and CGRP are related peptides (Fig. 1); second, there is evidence for the existence of both CGRP and ADM receptors in the lung (where CRLR, RAMP2 and RAMP3 are co-expressed); and finally, receptors have been described that are sensitive to both ADM and CGRP^{1,2,20–23}. Oocytes expressing RAMP2 and CRLR (plus CFTR) responded to ADM with large inward currents at concentrations of ADM₁₃₋₅₂ (residues 13–52 of ADM) as low as 0.3 nM, but CGRP produced only small responses (Fig. 8a). The

responses to ADM₁₋₅₂ and ADM₁₃₋₅₂ were similar. ADM₂₂₋₅₂ and CGRP₈₋₃₇ had no agonist activity (at 100 nM) but significantly antagonized the response to 10 nM ADM₁₃₋₅₂ (Fig. 8b). No responses were observed to either calcitonin or amylin at 300 nM (not shown).

Expression of RAMP2 failed to generate responses to ADM without CRLR (not shown). This may be because RAMP2 cannot interact with the *Xenopus* homologue of CRLR or because the *Xenopus* receptor expressed with RAMP2 does not recognize

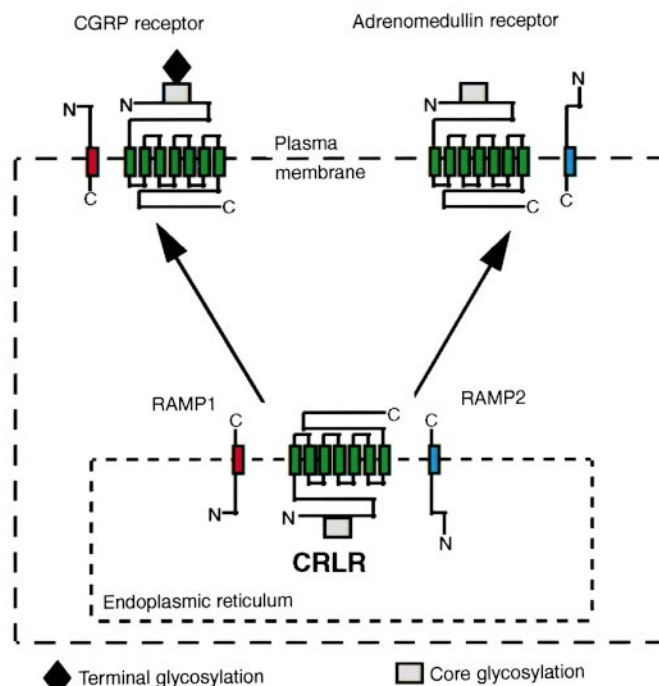


Figure 9 The role of RAMPs 1 and 2 and CRLR in generating CGRP or ADM receptors.

human ADM. In either case, the new response appeared to result entirely from exogenous RAMP2 and CRLR.

Expression of CRLR with RAMP2 in HEK293T cells generated new ADM-stimulated responses of cAMP (Fig. 8c) and specific binding to 125 I-labelled rat ADM (Table 1). The absolute levels of specific binding of 125 I-labelled rat ADM to co-transfected HEK293T cells and to the native ADM receptors in NG108-15 cells were similar. We constructed competition curves for binding of 125 I-labelled rat ADM to membrane preparations from both cell lines in the same set of experiments (Table 1). The half-maximal inhibitory concentrations (IC_{50} values) are similar in each case and are consistent with the pharmacology of the reconstituted ADM receptor in oocytes. Some interspecies variation might be expected as NG108-15 cells derive from a rat–mouse cell fusion³.

The receptor generated by expression of RAMP3 and CRLR is similar to the ADM receptor and is being studied.

Discussion

We have shown that the CRLR has two alternative pharmacological profiles that are conferred by the accessory proteins RAMP1 (producing the CGRP receptor) and RAMP2 (producing the ADM receptor). RAMPs control the transport and glycosylation of the CRLR (Fig. 9). The differences in pharmacology between RAMP1–CRLR and RAMP2–CRLR could be due to the differential glycosylation of CRLR or to the presence of the RAMP at the plasma membrane (or possibly both). The RAMPs and CRLR can be closely associated and are expressed together at the cell surface. RAMPs 1 and 2 are endogenous to at least three cell lines. Expression of CRLR (either from the native gene or by a recombinant route) with RAMPs would generate a pharmacology dependent on the proportion of CRLR that is transported by each RAMP. A 'pure' CGRP receptor or ADM receptor population might only exist in cells that express CRLR with a single RAMP. However, in practice there is some consensus regarding 'typical' CGRP receptor and ADM receptor pharmacological profiles, and these are represented by the receptors native to SK-N-MC and NG108-15 cells, respectively. The pharmacologies of CGRP and ADM receptors in tissues are less clear, but the responses can be complex as many cells and cell types contribute receptors.

Attempts to identify calcitonin and amylin receptors gave some results similar to ours. These two receptors have similar protein backbones²⁴ and attempts to clone the gene encoding the amylin receptor using a selective ligand resulted in the isolation of a cDNA encoding the calcitonin receptor²⁵. Production of amylin receptors from this cDNA was cell-line-specific (however, calcitonin receptors were always synthesized, with a variable proportion of receptors having amylin receptor pharmacology)²⁵. Our data might allow a molecular understanding of some of the complexities associated with the pharmacology of calcitonin and amylin as well as of CGRP and ADM.

The existence of RAMPs indicates a new mechanism whereby cells/tissues could change their responsiveness to different neuropeptides. Whether this is a regulatory event remains to be determined. There will be considerable interest in whether RAMPs regulate other receptor systems. Many receptors are not easily expressed in commonly used host cell lines: GABA_B, eotaxin (CKR3)²⁶, thrombin (PAR3) and olfactory receptors are just a few examples of such receptors. RAMPs may have a wider role, as they are expressed at relatively high levels in many tissues and can be expressed independently of CRLR.

We have now unambiguously identified the CGRP and ADM receptors, described the RAMP family of proteins, and presented a new mechanism for regulating GPCRs that may have physiological importance. □

Methods

Oocytes and expression cloning. We injected capped cRNA into stage V–VI

defolliculated oocytes¹⁴. Each oocyte received 50–100 ng CFTR RNA and 1–20 ng cloned RNAs or 100–400 ng SK-N-MC pool RNA and was tested for responses to CGRP after 3–14 days, under two-electrode voltage clamp. We used the endogenous oocyte β -adrenergic receptor to control for CFTR expression and to ensure that any increased responses were specific to CGRP. A pool of about 0.7 million clones produced a CFTR-mediated response to CGRP. Six further subdivisions were made, corresponding to pool sizes of ~50,000, ~20,000, 4,000 (gridded), 168, 40 and 2 clones respectively. Oocytes were superfused with a Ca^{2+} -free ND96 solution (96 mM NaCl, 2 mM KCl, 1 mM $MgCl_2$, 5 mM HEPES, pH 7.5 at 25 °C) at a flow rate of 2 ml min⁻¹. Recording electrodes had resistance of 0.5–1.0 M Ω when filled with 3 M KCl. Human ADM peptides (residues 1–52, 13–52 and 22–52) were obtained from Peninsula Labs. Amylin, calcitonin, CGRP₂, CGRP_{8–37} and VIP were obtained from Bachem and CGRP1 was synthesized at GlaxoWellcome.

Molecular biology. CRLR⁸ was modified to provide a consensus Kozak sequence as described¹². Myc and haemagglutinin (HA) epitope tags were used in-frame to the 5' end of cDNAs encoding RAMP1 and CRLR. In each case, we removed the native signal sequence and replaced it with that for CD33 (ref. 27) (Myc–CRLR) or T8 (HA–CRLR). Myc–RAMP1, Myc–CRLR and HA–CRLR were cloned into pcDNA3 (Invitrogen). All modified sequences were compared to 'native' sequences in the assays described and found to behave identically (data not shown). Human RDC1 was cloned by PCR from genomic DNA and canine RDC1 was provided by A. Clarke¹⁰.

Cell culture and cDNA transfection. HEK293T and Swiss3T3 cells were maintained in DMEM medium supplemented with 10% fetal calf serum (FCS) and 2 mM glutamine and maintained at 37 °C and 95% humidity. NG108-15 cells were cultured in DMEM/HAM F12 (1:1) supplemented with hypoxanthine–aminopterin–thymidine (HAT) (1 $\times$) and 10% FCS.

Stable Swiss 3T3 cells expressing CRLR and RAMP1 were generated by dual hygromycin and neomycin selection imposed on cells transfected with plasmids encoding both proteins and different selection markers. Individual lines were cloned by dilution.

5 μ g linearized DNA per plasmid was transfected for each T 75 cm² flask containing 80% confluent cells. Cells were left for 24 h before plating into 96-well plates and were then cultured for a further day. For cAMP determinations, the cells were washed with PBS and preincubated in DMEM medium containing 300 μ M IBMX (Sigma) for 30 min at 37 °C. CGRP or ADM was added for a further 30 min and the cells were washed with ice-cold PBS. cAMP levels were determined using scintillation proximity based assays (Amersham).

Radioligand binding, crosslinking analysis, SDS-PAGE and western blotting. HEK293T cells were collected after 2 days (NG108-15 cells were collected when confluent) into PBS. Cells were pelleted by centrifugation and homogenized in 50 mM HEPES-KOH, pH 7.6 (containing 15 μ g ml⁻¹ aprotinin, 0.25 μ g ml⁻¹ antipain, 0.25 μ g ml⁻¹ leupeptin, 0.1 μ g ml⁻¹ benzamide and 0.1 μ g ml⁻¹ bacitracin). After centrifugation at 500 g for 20 min at 4 °C, the supernatant was removed and centrifuged at 48,000g for 30 min at 4 °C. The final pellet was resuspended in homogenization buffer and the protein content measured. SK-N-MC and CRLR/RAMP1 double stable Swiss 3T3 cell lines were collected into PBS but subjected to the same membrane preparation in 50 mM HEPES-KOH, 1 mM EDTA, 100 μ M leupeptin, 25 μ g ml⁻¹ bacitracin, pH 7.6. Immediately before the first homogenization, 1 mM phenylmethylsulphonyl fluoride (PMSF) and 2 μ M pepstatin A were added. The final pellet was resuspended without PMSF or pepstatin A.

For the CGRP receptor assay, membranes (50 μ g) were incubated for 90 min at 25 °C in binding buffer (50 mM HEPES-KOH, 10 mM $MgCl_2$, 1 mM EDTA, pH 7.4), containing 30 pM 125 I-labelled CGRP1 (Amersham) in a total volume of 200 μ l. For the ADM receptor assay, membranes (10–20 μ g) were incubated for 30 min at 4 °C in the same buffer with 100 pM 125 I-labelled rat ADM (Amersham) in 200 μ l. Incubation for both assays was terminated by rapid filtration through GF/C filters soaked in 0.1% polyethylenimine using a Tomtek cell harvester. Nonspecific binding was determined using a final concentration of either 1 μ M CGRP or 1 μ M ADM_{13–52}.

For crosslinking analysis, cells were transfected with cDNA and, after two days, detached and incubated with 125 I-labelled CGRP1 with or without unlabelled CGRP and crosslinked at 4 °C with either BS³ or DSS (disuccinimidyl suberate)¹⁷ (not shown but similar results) (reagents from Pierce). The cells were centrifuged and then extracted into sample buffer¹⁷ and subjected to

SDS–PAGE and autoradiography. For other experiments, plasma-membrane-containing P2 particulate fractions were prepared from transfected cell pastes that had been stored at -80°C following collection. Membrane protein (75 μg) was subjected to 10% SDS–PAGE. Epitope tags were visualized by immunoblotting with anti-Myc or anti-HA monoclonal antibodies and developed using enhanced chemiluminescence (Amersham).

FACS analysis. HEK293T cells were transiently transfected with cDNA as described, collected 2 days later and washed three times in PBS. The cells were resuspended in DMEM and incubated with the primary antibody, 9E10 (anti-c-Myc), diluted 1:30 for 15 min. Following three further washes, the secondary antibody (sheep anti-mouse Fab₂ coupled with fluorescein isothiocyanate) diluted 1:30 was incubated for 30 min in the dark. For permeabilized cells, the Fix and Perm kit (Caltag) was used. FACS analysis was performed on an EPICS Elite (Coulter, Florida). 10,000 cells were sorted in each experiment.

Deglycosylation was carried out according to the supplier's protocols (Boehringer) before SDS–PAGE/immunoblotting.

Received 20 November 1997; accepted 2 April 1998.

1. Poyner, D. R. Calcitonin gene-related peptide: multiple actions, multiple receptors. *Pharmacol. Ther.* **56**, 23–51 (1992).
2. Muff, R., Born, W. & Fischer, J. A. Calcitonin, calcitonin gene related peptide, adrenomedullin and amylin: homologous peptides, separate receptors and overlapping biological actions. *Eur. J. Endocrinol.* **133**, 17–20 (1995).
3. Zimmermann, U. *et al.* Identification of adrenomedullin receptors in cultured rat astrocytes and in neuroblastoma $\times$ glioma hybrid cells (NG108-15). *Brain Res.* **724**, 238–245 (1996).
4. Wang, X. *et al.* Discovery of adrenomedullin in rat ischaemic cortex and evidence for its role in exacerbating focal brain ischaemic damage. *Proc. Natl Acad. Sci. USA* **92**, 11480–11484 (1995).
5. Lin, H. Y. *et al.* Expression cloning of an adenylate cyclase-coupled calcitonin receptor. *Science* **254**, 1022–1024 (1991).
6. Njuki, F. *et al.* A new calcitonin-receptor-like sequence in rat pulmonary blood vessels. *Clin. Sci.* **85**, 385–388 (1993).
7. Chang, C.-P., Pearse, R. V., O'Connell, S. & Rosenfeld, M. G. Identification of a seven transmembrane helix receptor for corticotropin-releasing factor and sauvagine in mammalian brain. *Neuron* **11**, 1187–1195 (1993).
8. Fluhmann, B., Muff, R., Hunziker, W., Fischer, J. A. & Born, W. A human orphan calcitonin receptor-like structure. *Biochem. Biophys. Res. Commun.* **206**, 341–347 (1995).
9. Kapas, S., Catt, K. J. & Clark, A. J. Cloning and expression of cDNA encoding a rat adrenomedullin receptor. *J. Biol. Chem.* **270**, 25344–25347 (1995).

10. Kapas, S. & Clark, A. J. L. Identification of an orphan receptor gene as a type 1 calcitonin gene-related receptor. *Biochem. Biophys. Res. Commun.* **217**, 832–838 (1995).
11. Luebke, A. E., Dahl, G. P., Roos, B. A. & Dickerson, I. M. Identification of a protein that confers calcitonin gene-related peptide responsiveness to oocytes by using a cystic fibrosis transmembrane conductance regulator assay. *Proc. Natl Acad. Sci. USA* **93**, 3455–3460 (1996).
12. Aiyar, N. *et al.* A cDNA encoding the calcitonin gene-related peptide type 1 receptor. *J. Biol. Chem.* **271**, 11325–11329 (1996).
13. Zimmermann, U. *et al.* Calcitonin gene-related peptide (CGRP) receptors are linked to cyclic adenosine monophosphate production in SK-N-MC human neuroblastoma cells. *Neurosci. Lett.* **119**, 195–198 (1990).
14. Uezono, Y. *et al.* Receptors that couple to 2 classes of G proteins increase cAMP and activate CFTR expressed in *Xenopus* oocytes. *Recept. Chan.* **1**, 233–241 (1993).
15. Guillemare, E., Lazdunski, M. & Honore, E. CGRP-induced activation of K_{ATP} channels in follicular *Xenopus* oocytes. *Pflügers Arch.* **428**, 604–609 (1994).
16. Han, Z.-Q. *et al.* The interaction of CGRP and adrenomedullin with a receptor expressed in the rat pulmonary vascular endothelium. *J. Mol. Endocrinol.* **18**, 267–272 (1997).
17. Stangl, D., Born, W. & Fischer, J. A. Characterisation and photoaffinity labeling of a Calcitonin Gene-Related Peptide receptor solubilised from human cerebellum. *Biochemistry* **30**, 8605–8611 (1991).
18. Foord, S. M. & Craig, R. K. Isolation and characterisation of a human calcitonin-gene-related-peptide receptor. *Eur. J. Biochem.* **170**, 373–379 (1987).
19. Teasdale, R. D. & Jackson, M. R. Signal mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus. *Annu. Rev. Cell Dev. Biol.* **12**, 27–54 (1996).
20. Zimmermann, U. *et al.* Adrenomedullin and calcitonin gene-related peptide interact with the same receptor in cultured human neuroblastoma SK-N-MC cells. *Peptides* **16**, 421–424 (1995).
21. Owji, A. A. *et al.* Characterisation and molecular identification of adrenomedullin binding sites in the rat spinal cord: a comparison with calcitonin gene-related peptide receptors. *J. Neurochem.* **67**, 2172–2179 (1996).
22. Osajima, A. *et al.* Adrenomedullin-sensitive receptors are preferentially expressed in cultured rat mesangial cells. *Eur. J. Pharmacol.* **315**, 319–325 (1996).
23. Kato, J., Kitamura, K., Kangawa, K. & Eto, T. Receptors for adrenomedullin in human vascular endothelial cells. *Eur. J. Pharmacol.* **289**, 383–385 (1995).
24. Perry, K. J. *et al.* Characterization of amylin and calcitonin receptor binding in the mouse alpha-thyroid-stimulating hormone thyrotroph cell line. *Endocrinology* **138**, 3486–3496 (1997).
25. Chen, W. J. *et al.* Expression cloning and receptor pharmacology of human calcitonin receptors from MCF-7 cells and their relationship to amylin receptors. *Mol. Pharmacol.* **52**, 1164–1175 (1997).
26. Ponath, P. D. *et al.* Molecular cloning and characterization of a human eotaxin receptor expressed selectively on eosinophils. *J. Exp. Med.* **183**, 2437–2448 (1996).
27. Simmons, D. & Seed, B. Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. *J. Immunol.* **141**, 2797–2800 (1988).

Acknowledgements. We thank D. Cousins for probing the MTN blots; E. Murrison for help with cloning RAMP2; S. Brown for isolation of the CRLR/RAMP1 cell lines; J. M. Rommens and L. C. Tsui for CFTR; W. Born and J. Fischer for CRLR and for discussions; and members of the RSU for their support.

Correspondence and requests for materials should be addressed to S.M.F. (e-mail: smf3746@ggr.co.uk). The cDNA sequences for human RAMPs 1, 2 and 3 have been submitted to the EMBL database under the accession numbers AJ001014, AJ001015 and AJ001016, respectively.

Reflected infrared spectrum of a massive protostar in Orion

Jun-Ichi Morino*, Takuya Yamashita†, Tetsuo Hasegawa* & Takenori Nakano‡

* Institute of Astronomy, The University of Tokyo, Osawa, Mitaka, Tokyo 181-8588, Japan

† National Astronomical Observatory, Osawa, Mitaka, Tokyo 181-8588, Japan

‡ Nobeyama Radio Observatory, National Astronomical Observatory, Minamimaki, Minamisaku, Nagano 384-1305, Japan

The infrared source IRC2 (ref. 1) in the star-forming region Orion-KL is generally believed to contain a massive and very young star². Its nature and evolutionary status, however, are difficult to determine because it is hidden from direct view by a dense disk-like envelope of gas and dust. Here we report observations of infrared radiation (at a wavelength of about 2 μm) that has escaped the surrounding dust in the polar direction, perpendicular to the plane of the disk, and then been reflected towards us by dust farther away from the star. The reflected spectrum contains absorption lines of neutral metallic atoms and carbon monoxide, which we interpret as indicating a source temperature of about 4,500 K. But, given the luminosity of the source, its radius must be at least 300 solar radii—too large to be attained with the modest gas-accretion rates in existing theories of massive-star formation. Whether the infrared radiation is coming from the protostar itself or the self-luminous accretion disk around it, the accretion rate must be around $(5\text{--}15) \times 10^{-3}$ solar masses per year, at least two orders of magnitude greater than is commonly assumed in models of star formation.

Figure 1 shows a K'-band (2.15- μm) image of the Orion-KL region. IRC2, the target object of the present work, is embedded in the narrow dust lane that separates the bipolar nebulosity, and is visible only at longer wavelengths. In the standard picture² that emerged in the 1980s, IRC2 is postulated to be a single massive star in formation. Its large bolometric luminosity, estimated as $(0.4\text{--}2) \times 10^5 L_{\odot}$ (ref. 3), dominates the region shown in Fig. 1 along with the Becklin-Neugebauer object (BN). IRC2 exhibits intense SiO maser emission⁴, which is generally associated with red giant stars but is rare in star-forming regions^{5,6}. It is encircled by a dense disklike gas^{7,8}, and an energetic mass outflow emerges from it forming a bipolar outflow perpendicular to the disk^{9,10}. The butterfly-shaped nebulosity in Fig. 1 is due to reflection of the 2- μm light from IRC2 that leaks through the cavities created by the outflow¹¹.

Figure 2 shows the spectrum of the light integrated over the brightest part of the lobe east of IRC2, along with a comparison spectrum of a red supergiant star HD12014 (K0Ib). The spectrum of the lobe is characterized by a very red continuum with emission and absorption lines. The emission lines are the H $\text{Br}\gamma$ recombination line from the Orion nebula in the foreground and the H_2 rotation-vibration lines from the gas shocked by the outflow¹². The absorption lines are assigned to neutral Na and Ca atoms and to the $\Delta v = 2$ rotation-vibration bands of carbon monoxide (CO), which are commonly observed in photospheres of late-type (low-temperature) stars as is visible in the spectrum of HD12014. The infrared nebula is highly polarized, with the electric vectors in the eastern lobe being concentric about the position of IRC2 (ref. 11); the CO absorption depth does not vary significantly over the extent of the lobe covered by the slit. These observations indicate that the absorption lines are not of interstellar origin but are most probably formed at IRC2 itself or in its vicinity. This gives us an opportunity to analyse the light from the massive protostar for the first time.

The absorption strengths of the atomic lines and the CO bands are established diagnostics of the photospheric temperature and surface gravity of late-type stars^{13,14}. But their application to the IRC2 spectrum does not give a consistent answer. The temperature deduced from the relatively weak atomic lines ($\sim 5,500$ K) contradicts that required by the prominent CO absorption ($\leq 5,000$ K). The shape of the CO bands is also different: the depth of the CO absorption at the $v = 0 \rightarrow 2$ bandhead relative to that at the flat part in the band at a longer wavelength is 1.6 for IRC2, distinctly smaller than 3.3 in HD12014 and other late-type stars (4–5 in giants and 3–4 in supergiants observed with the same equipment). This difference indicates that the intrinsic width of the saturated individual lines in the CO bands is much larger in IRC2, possibly reflecting a very low surface gravity or an excess turbulent motion in the gas. This in turn predicts a very deep CO absorption in IRC2, which is apparently not observed.

The above inconsistency calls for the introduction of a secondary featureless continuum component to the IRC2 spectrum. For example, dust emission from the disk surrounding the protostar might add some flux density to the continuum. Indeed, such a

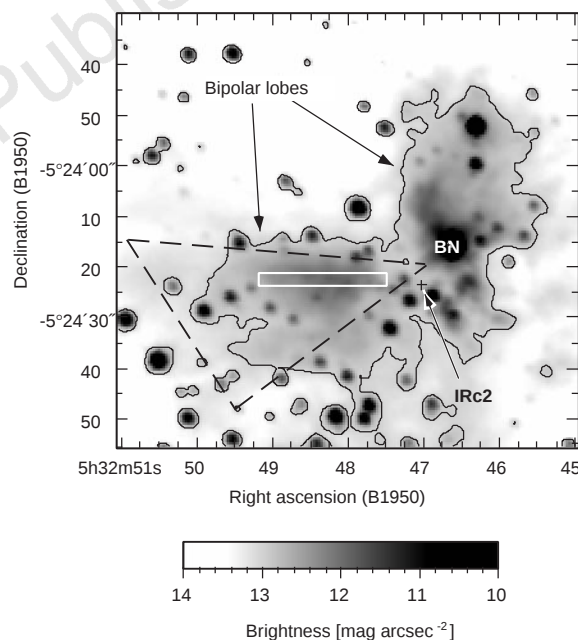


Figure 1 A K'-band ($\lambda_{\text{centre}} = 2.15 \mu\text{m}$) image of the Orion-KL region taken in 1996 December. The image was obtained with the near-infrared camera/spectrometer system OASIS mounted on the 1.9-m telescope at the Okayama Astrophysical Observatory (OAO), National Astronomical Observatory of Japan. A square field of $\sim 4 \text{ arcmin} \times 4 \text{ arcmin}$ was covered by a 256 pixel $\times$ 256 pixel NICMOS3 detector array with a scale of 0.94 arcsec/pixel. Only part of the frame is shown. The brightest point source near the centre is the Becklin-Neugebauer (BN) object, which is known to be a young B-type star with a bolometric luminosity of $\sim 10^4 L_{\odot}$ (ref. 3). The contours are drawn at 13 mag arcsec⁻². The white rectangle indicates the area integrated to make the spectrum of the reflected light shown in Fig. 2. We estimate the flux density of IRC2, the illuminating source, by integrating the brightness of the nebula within the dashed triangle. The standard picture of IRC2 (ref. 2) adopted in this paper is challenged by recent observations that reveal a 0.7'' spatial offset of the 10- μm source from the SiO maser/radio source^{25,26} and resolve the 10- μm source into three knots at 3.8 μm (ref. 27). However, the linear polarization of the 3.8- μm knots leaves room for the standard picture to survive; in other words, the knots might not be self-luminous but be illuminated by the luminosity source associated with the SiO maser. Even if the luminosity source is not a single protostar but a group of such objects, a large fraction of the total luminosity might arise from the most massive object. With all these reservations, we adopt the standard picture in this Letter and use the name IRC2 for the most massive object that dominates the activity in this region.

component (infrared excess) is frequently found in the spectra of young stars¹⁵. If, for example, we assume that the featureless component contributes half of the continuum flux density of IRC2 and dilutes the absorption lines by a factor of two, the equivalent widths of the atomic lines and the CO index (at the bandhead) become quite consistent with those for HD12014, which has an effective temperature of 4,500 K. Considering the uncertainties in the contribution of the secondary component and in the measured indices, we conclude that the absorption features in IRC2

are formed at a temperature of 3,000–5,500 K; the most probable temperature is 4,500 K.

Another important clue to the nature of the object comes from its intrinsic near-infrared luminosity estimated from the surface brightness of the reflection nebula. A part of the thermal emission originating from the object escapes through the bipolar cavities and is scattered at the cavity wall. At the same time the light is extinguished and reddened along its path from the object toward us. To derive the intrinsic luminosity we follow each of the above processes. First we estimate the K'-band flux density of the reflected light in the eastern lobe to be ~ 6.5 mag by integrating over the triangular area shown in Fig. 1. Contributions from the point sources and the foreground free-free emission ($\sim 40\%$) have been subtracted. If we assume an opening angle of 45° for the bipolar cavity and an albedo, the ratio of the scattered light to the extinguished (scattered plus absorbed) light, of 0.22 (ref. 16), the flux density of the eastern lobe should be $\sim 3.3\%$ of that of the illuminator. The very red continuum in Fig. 2 indicates that the light is heavily extinguished. We take the intrinsic colour temperature of the illuminator to be 2,500 K, lower than the effective temperature derived above, considering a possible contribution from the redder featureless continuum. Adopting a reddening law of the form $A_\lambda = A_K \times (\lambda/2.15 \mu\text{m})^{-p}$ with $p = 1.7$ (ref. 17), the slope of the IRC2 spectrum in Fig. 2 corresponds to an extinction of 3.9 mag in the K' band. We neglect possible bluing caused by the scattering in view of uncertainties in scattering processes. This, together with the assumed low intrinsic colour temperature, leads us to the minimum extinction. After correcting for this extinction and the distance modulus of 8.3 mag (450 pc) (ref. 2), we get an intrinsic absolute magnitude M_K of ≤ -9.4 mag. For comparison, a 3,000–5,500-K blackbody with a bolometric luminosity of $10^5 L_\odot$ would have an absolute magnitude M_K of -10.5 to -9.1 mag. The broad agreement of the derived lower limit with these values supports our assumption that the illuminator of the infrared nebula and the source of the bolometric luminosity are identical.

What is the nature of the source of the luminosity? The presence of SiO maser emission and the infrared spectrum similar to a red supergiant star might raise a possibility that an evolved star that happens to be embedded in this region contributes much of the near-infrared light. But this is unlikely. The SiO maser in IRC2 is stable over time and exhibits an ordered kinematics suggestive of a rotating and expanding 80-AU disk¹⁸. This behaviour differs sharply from that of evolved stars in which maser spots appear rather randomly at a few stellar radii and the line profile changes drastically with time⁶. In addition, SiO maser sources are known to be associated with two other regions of massive star formation, namely W51 IRS2 and Sgr B2 MD5 (refs 5, 19). Thus the luminosity source is most probably the protostar. A forming star grows by accretion of gas through a disk surrounding it, in which a fraction of the gravitational energy is released. The central protostar evolves towards the main sequence through quasi-static contraction by radiating the internal thermal energy. So, two regions are potentially important in emitting the observed near infrared light, namely the surface (photosphere) of the protostar, and the surface of the accretion disk.

If the protostar itself is the source of much of the luminosity, its radius should be large. Assuming that the blackbody radiation of a 4,500-K photosphere contributes half of the K'-band flux estimated above, we derive its radius as being larger than $\sim 300 R_\odot$ and a corresponding bolometric luminosity being $\geq 4 \times 10^4 L_\odot$. This set of parameters locates IRC2 high up in the Hertzsprung–Russell diagram of stellar evolution, close to red supergiant stars that have evolved off the main sequence. A massive star can experience a common state characterized by a large radius and a low surface temperature at two extreme phases in its life. This raises a challenging question to the theory of early stellar evolution: can an accreting protostar be so big? The modest mass accretion rates of

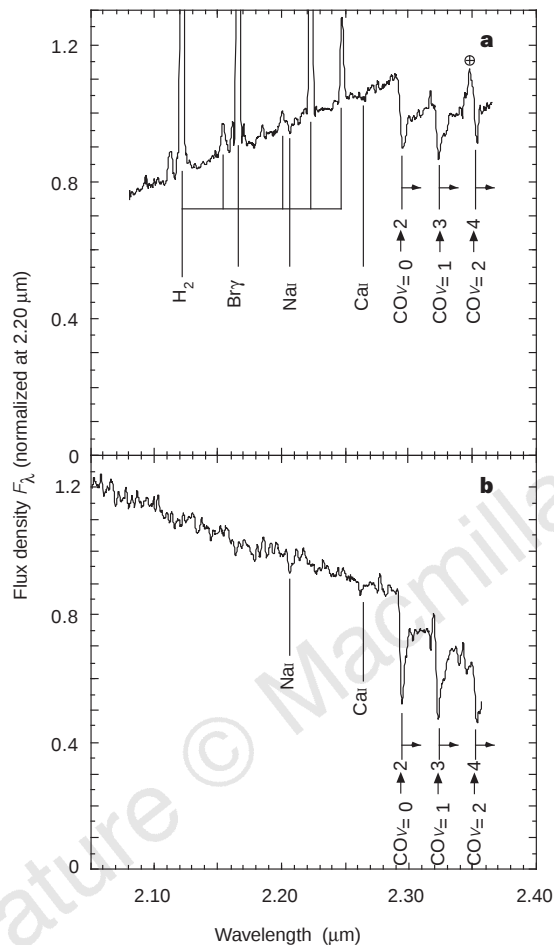


Figure 2 The K-band spectra taken with OASIS on the OAO 1.9-m telescope in grating spectroscopy mode. **a**, The diffuse emission of Orion-KL IRC2, eastern lobe, taken in 1997 January. The detector array covers a wavelength range of 2.08–2.42 μm at a resolution of $\lambda/\Delta\lambda \approx 1000$. A 4-arcmin-long and 2.4-arcsec-wide slit was placed east-west on the eastern lobe. The spectra were integrated over a length of 27 arcsec along the slit shown by the white rectangle in Fig. 1. **b**, HD12014, a red supergiant (K0Ib) star, taken in 1996 October, for comparison. The continuum in the IRC2 spectrum is extremely red. Assignments in both panels are given for the spectral features including the neutral Na ($\lambda 2.337 \mu\text{m}$) and Ca ($\lambda 2.263 \mu\text{m}$) atoms, and for the $\Delta v = 2$ rotation-vibration bands of CO ($\lambda > 2.295 \mu\text{m}$). The measured absorption indices (equivalent widths) of the atomic lines are $W(\text{CaI}) = 1.7 \pm 0.3 \text{ \AA}$ and $W(\text{NaI}) = 1.1 \pm 0.2 \text{ \AA}$, and the CO index (the depth of the CO absorption at the $v = 0 \rightarrow 2$ bandhead measured against the continuum linearly extrapolated from the shorter wavelength) is 0.19 ± 0.02 . For comparison, the corresponding values for HD12014 are $W(\text{CaI}) = 2.7 \pm 0.7 \text{ \AA}$ and $W(\text{NaI}) = 2.7 \pm 0.6 \text{ \AA}$, and the CO index is 0.40 ± 0.04 . The shape of the CO band is somewhat different in IRC2 and in HD12014. The ratio of the depth at the $v = 0 \rightarrow 2$ bandhead relative to that at the flat part in the band near $\lambda 2.31 \mu\text{m}$ is 1.6 for IRC2. Corresponding ratios are 3.3 in HD12014, 3–4 in G8–M0 supergiant stars, and 4–5 in G0–K5 giant stars observed with the same equipment. This difference is expected if the intrinsic width of the individual lines that compose the CO bands gets larger in the order giants < supergiants < IRC2.

10^{-4} – $10^{-5} M_{\odot} \text{yr}^{-1}$ commonly assumed for a massive protostar cannot make the star so big because its structural evolution (that is, its contraction) toward the main sequence is faster than injection of its internal energy through mass accretion²⁰. Nakano *et al.* (manuscript in preparation) made a simplified model with the polytrope approximation, and drew evolutionary tracks in the Hertzsprung–Russell diagram of protostars steadily accreting at larger rates. For an assumed polytropic index N of 4.0, the model reproduces IRC2 as a $20 M_{\odot}$ protostar with $300 R_{\odot}$ accreting at a rate as large as $\dot{M} \geq 5 \times 10^{-3} M_{\odot} \text{yr}^{-1}$. This implies a short accretion timescale of $\leq 4 \times 10^3 \text{yr}$, which is consistent with the short dynamical age ($1.5 \times 10^3 \text{yr}$) (ref. 21) of the molecular outflow that is suggested to be contemporary with mass accretion²².

Alternatively, if the accretion disk around a massive protostar is the major source of the infrared radiation, it would be the first such object to which we could apply the spectroscopic diagnosis. We assume that the central star has a mass of 10 – $25 M_{\odot}$ and a relatively small radius to keep the stellar contribution to the observed spectrum small. Gravitational energy released within the accretion disk is the source of the near-infrared luminosity. In the outer part of the disk, dust opacity hides gaseous absorption features to produce featureless continuum emission, which can explain the secondary component that we introduced to understand the IRC2 spectrum. In the inner part, the temperature is high enough for the dust grains to evaporate, and gaseous absorption lines appear²³. Indeed, a class of low-mass pre-main-sequence stars called FU Orionis-type stars exhibit deep absorption features of CO and H_2O in the near-infrared spectrum²⁴. If half of the K'-band luminosity of IRC2 comes from a scaled-up version of this dust-free part of accretion disks, the mass accretion rate must be at least $(6\text{--}15) \times 10^{-3} M_{\odot} \text{yr}^{-1}$. At these accretion rates, the bolometric luminosity of $10^3 L_{\odot}$ can be released if the radius of the inner edge of the disk is $\sim 20 R_{\odot}$. This poses a potential difficulty: the large accretion rate should have made the protostar much larger, as we have seen above, though this difficulty can be avoided if the accretion rate varies with time.

We might possibly be witnessing a short episode of active accretion around the small central star, or the central protostar might be large and contribute significantly to the observed spectral features. Follow-up observations of IRC2 with higher spectral resolution and searches for similar objects are strongly desired. Velocity-resolved spectra would clarify whether the star or the disk contributes more to the near-infrared light, uncovering the gas motion in the immediate vicinity of massive protostars. □

Received 5 August 1997; accepted 3 March 1998.

- Rieke, G. H., Low, F. J. & Kleinmann, D. E. High-resolution maps of the Kleinmann–Low nebula in Orion. *Astrophys. J.* **186**, L7–L11 (1973).
- Genzel, R. & Stutzki, J. The Orion molecular cloud and star-forming region. *Annu. Rev. Astron. Astrophys.* **27**, 41–85 (1989).
- Wynn-Williams, C. G., Genzel, R., Becklin, E. E. & Downes, D. The Kleinmann–Low nebula: an infrared cavity. *Astrophys. J.* **281**, 172–183 (1984).
- Snyder, L. E. & Buhl, D. Detection of possible maser emission near 3.48 millimeters from an unidentified molecular species in Orion. *Astrophys. J.* **189**, L31–L33 (1974).
- Hasegawa, T. *et al.* in *Masers, Molecules and Mass Outflows in Star Forming Regions* (ed. Haschick, A. D.) 275–287 (Haystack Observatory, Massachusetts, 1986).
- Elitzur, M. *Astronomical Masers* (Kluwer, Dordrecht, 1992).
- Hasegawa, T. *et al.* CS around Orion-KL: a large rotating disk. *Astrophys. J.* **283**, 117–122 (1984).
- Wright, M. C. H., Plambeck, R. L. & Wilner, D. J. A multilane aperture synthesis study of Orion-KL. *Astrophys. J.* **469**, 216–237 (1996).
- Erickson, N. R. *et al.* Detection of bipolar CO outflow in Orion. *Astrophys. J.* **261**, L103–L107 (1982).
- Wright, M. C. H., Plambeck, R. L., Vogel, S. N., Ho, P. T. P. & Welch, W. J. Source of the high-velocity molecular flow in Orion. *Astrophys. J.* **267**, L41–L45 (1983).
- Minchin, N. R. *et al.* Near-infrared imaging polarimetry of bipolar nebulae—I. The BN-KL region of OMZC-1. *Mon. Not. R. Astron. Soc.* **248**, 715–729 (1991).
- Beckwith, S., Persson, S. E., Neugebauer, G. & Becklin, E. E. Observations of the molecular hydrogen emission from the Orion nebula. *Astrophys. J.* **223**, 464–470 (1978).
- Kleinmann, S. G. & Hall, D. N. B. Spectra of late-type standard stars in the region 2.0–2.5 microns. *Astrophys. J. Suppl. Ser.* **62**, 501–517 (1986).
- Ali, B., Carr, J. S., DePoy, D. L., Frogel, J. A. & Sellgren, K. Medium-resolution near-infrared (2.15–2.35 μm) spectroscopy of late-type main-sequence stars. *Astron. J.* **110**, 2415–2424 (1995).
- Greene, T. P. & Lada, C. J. Near-infrared spectra and the evolutionary status of young stellar objects: results of a 1.1–2.4 μm survey. *Astron. J.* **112**, 2184–2221 (1996).

- Draine, B. T. & Lee, H. M. Optical properties of interstellar graphite and silicate grains. *Astrophys. J.* **285**, 89–108 (1984).
- Mathis, J. S. Interstellar dust and extinction. *Annu. Rev. Astron. Astrophys.* **28**, 37–70 (1990).
- Plambeck, R. L., Wright, M. C. H. & Carlstrom, J. E. Velocity structure of the Orion-IRC2 SiO maser: evidence for an 80 AU diameter circumstellar disk. *Astrophys. J.* **348**, L65–L68 (1990).
- Morita, K.-I., Hasegawa, T., Ukita, N., Okumura, S. K. & Ishiguro, M. Accurate positions of SiO masers in active star-forming regions: Orion-KL, W51-IRS2, and Sagittarius-B2 MD5. *Publ. Astron. Soc. Jpn* **44**, 373–380 (1992).
- Palla, F. & Stahler, S. W. The pre-main sequence evolution of intermediate-mass stars. *Astrophys. J.* **418**, 414–425 (1993).
- Snell, R. L., Scoville, N. Z., Sanders, D. B. & Erickson, N. R. High-velocity molecular jets. *Astrophys. J.* **284**, 176–193 (1984).
- Shu, F. *et al.* Magnetocentrally driven flows from young stars and disks. I. A generalized model. *Astrophys. J.* **429**, 781–796 (1994).
- Calvet, N., Hartmann, L. & Kenyon, S. J. On the near-infrared spectrum of FU Orionis. *Astrophys. J.* **383**, 752–756 (1991).
- Hartmann, L. & Kenyon, S. J. The FU Orionis phenomenon. *Annu. Rev. Astron. Astrophys.* **34**, 207–240 (1996).
- Gezari, D. Y. Mid-infrared imaging of Orion BN/KL: astrometry of IRC2 and the SiO maser. *Astrophys. J.* **396**, L43–L47 (1992).
- Menten, K. M. & Reid, M. J. What is powering the Orion Kleinmann–Low infrared nebula? *Astrophys. J.* **445**, L157–L160 (1995).
- Dougados, C., Léna, P., Ridgway, S. T., Christou, J. C. & Probst, R. G. Near-infrared imaging of the Becklin–Neugebauer-IRC2 region in Orion with subarcsecond resolution. *Astrophys. J.* **406**, 112–121 (1993).

Acknowledgements. We thank S. Okumura and E. Nishihara for their kind help during the observations. J.-I.M. was supported by research fellowships of the Japan Society for the Promotion of Science for Young Scientists.

Correspondence and requests for materials should be addressed to J.-I.M. (e-mail: morino@mtk.iaa.s.u.tokyo.ac.jp).

Relation to solar activity of intense aurorae in sunlight and darkness

Patrick T. Newell, Ching-I. Meng & Simon Wing

Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland 20723-6099, USA

The oldest documented^{1,2} relationship between the number of sunspots (the solar cycle) and terrestrial effects is the increased frequency of aurorae in the period immediately after the solar maximum (the peak of the number of sunspots). This correlation is, however, based only on observations of the relatively rare events of 'great aurorae', which are those that reach mid-latitudes or lower. The overwhelming majority of intense aurorae, and therefore most of the energy put into the ionosphere, occurs at high latitudes, where aurorae appear nightly. Here we report the global frequency of aurorae as a function of solar cycle, determined by data from the US Air Force Defense Meteorological Satellite Program. We find that, contrary to expectations, the total number of intense aurorae is uncorrelated with solar activity in darkness, and is negatively correlated with solar activity in sunlit conditions. These findings imply a causal relationship between aurorae and ionospheric conductivity (the latter is maximal at solar maximum) and therefore indicate that the occurrence of intense aurorae is a discharge phenomenon, similar to lightning.

A few times a year an aurora expands to reach mid-latitude regions, more than 30° below the magnetic poles. Although there are several long-term records of these great auroral events, as observed by early, dedicated scientists at isolated points around the world, it is only a matter of common assumption whether the global frequency of aurorae, which are common at high latitudes, has a similar solar-cycle variation. Because relatively intense aurorae (we will use $5 \text{ erg cm}^{-2} \text{ s}^{-1}$ as a threshold criterion) occur almost every night over a wide region at high latitudes, the great auroral storms that happen a few times a year contribute only negligibly to the total energy input into the upper atmosphere. Because of the importance

of the auroral energy flux input to upper-atmospheric high-latitude chemistry³, and because of the intrinsic interest in understanding the physics of auroral arc formation, we have studied the effects of the solar cycle on the global frequency of intense aurorae. Specifically, based on recent research supporting the idea that ionospheric conductivity plays an important role in the formation of intense auroral arcs^{4,5}, we test the hypothesis that intense aurorae under sunlit conditions should be rarer at solar maxima, when ionospheric conductivity is higher.

Probably the only way to study intense auroral events on a global scale in both darkness and sunlight over the period of a solar cycle is to use the electron detectors on board satellites of the Air Force Defense Meteorological Satellite Program (DMSP). Since the end of 1983 there have been at least two such satellites operational and monitored continuously at all times, each equipped with essentially identical electron detectors⁶, allowing global coverage with unmatched longevity. These detectors measure the precipitating electron acceleration events that cause intense aurorae⁷. The procedure for the automated identification of such events has been discussed previously⁵. Electron acceleration events are characterized by 'mono-energetic' electron spectra, and their association with aurorae was established by direct rocket shots into visible aurorae⁷. We have measured the frequency of intense aurorae integrated over latitude from 60° to 80° magnetic latitude (MLAT) and in magnetic local time from 18:00 to 24:00 MLT. The magnetic coordinate system is similar to the more familiar geographic coordinates except that the magnetic pole replaces the geographic pole. The

latitudinal range of integration covers the main auroral zone; the local time integration covers the region of most intense discrete aurorae, but is limited to pre-midnight, partly because DMSP satellite coverage in some years was scanty after midnight.

Our results were constructed in two stages. First, the probability, $P(\text{MLAT}, \text{MLT})$ of observing intense electron acceleration events ($>5 \text{ erg cm}^{-2} \text{ s}^{-1}$ with a mono-energetic peak) was calculated for each $0.5^\circ \times 0.5 \text{ h}$ bin, based on the number of spectra fitting the criteria for an acceleration event divided by the total number of satellite spectra taken within that bin. The average surface 'area' covered by intense arcs was then calculated from $\int_{60}^{80} \int_{18}^{24} P(\text{MLT}, \text{MLAT}) \text{ d}(\text{MLAT}) \text{ d}(\text{MLT})$. The numbers thus arrived at are a measure of the time-averaged surface area covered with intense discrete aurorae.

Figure 1 shows the variation of intense auroral events over the course of a solar cycle. Figure 1a is restricted to measurements taken under sunlit conditions (solar zenith angle of 85° or less), whereas Fig. 1b included only measurements taken under dark conditions (solar zenith angle of 110° or larger). Because of the local time of the measurements (18:00 to 24:00 MLT), the sunlit measurements occur during local summer months (data from both hemispheres are combined, depending on the appropriate local solar zenith angle). For reference the solar flux at 10.7 cm (F10.7 number) is overlaid to show the variation in solar cycle. Although 10.7 cm is not an ionizing wavelength of solar emission, it is often used as a proxy for the level of solar flux including ultraviolet emissions, and has the advantage of long-term availability. Moreover, it has been shown

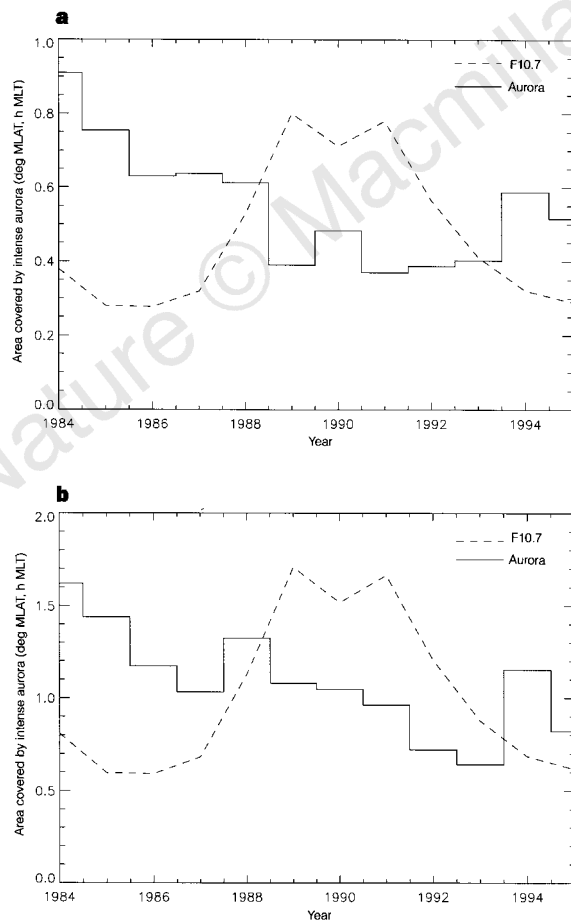


Figure 1 Mean surface area covered by intense ($\geq 5 \text{ erg cm}^{-2} \text{ s}^{-1}$) aurorae over a 12-year period. For comparison, the F10.7 number, which correlates well with solar cycle and is often used as a proxy for solar activity, is also plotted (broken line). **a**, Locally sunlit conditions (solar zenith angle $\leq 85^\circ$). **b**, Conditions of local darkness (solar zenith angle $\geq 110^\circ$).

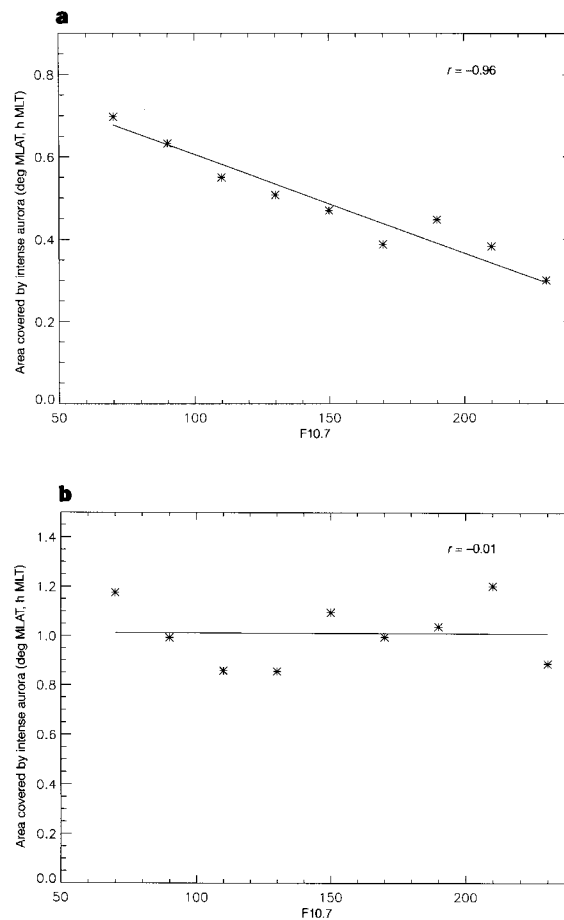


Figure 2 Mean surface area covered by intense auroral arcs as an explicit function of F10.7 daily value, which is a proxy for ionizing solar radiation. **a**, Sunlit conditions (solar zenith angle $\leq 85^\circ$). **b**, Local conditions of darkness (solar zenith angle $\geq 110^\circ$).

empirically that the F10.7 number correlates well with ionospheric conductivity⁸.

The sunlit (hence summer) hemisphere measurements seem to show a relatively clear solar cycle variation, with a pronounced minimum in auroral frequency around the solar maximum (about 1990). Under conditions of darkness, no clear solar cycle trend exists, or at least none that is clear enough to extract from the available 12-year stretch of data. Because most intense aurorae occur in darkness (both because they are more frequent when the ionosphere is not sunlit⁴ and because the local time of most intense aurorae (18:00 to 24:00 MLT) is in darkness for most of the year), the net effect is that solar cycle plays a relatively minor role in total precipitating energy-flux input into the ionosphere from intense aurorae. It will be helpful if the DMSP program continues for one or more additional solar cycles to help verify these results.

Although the yearly averages are of intrinsic interest, this approach is not the best for determining whether ionizing solar radiation is involved directly in suppressing the formation of intense aurorae. Over the course of a year both the frequency of aurorae and the F10.7 number can undergo substantial variations. It is therefore more useful to sort aurorae directly according to the 'instantaneous' (daily) value of F10.7 at the time of observation. The results of this approach are shown in Fig. 2. Each data point represents the time-averaged occurrence rate of aurora over the 12-year period 1984–1995 such that the daily F10.7 number was within the appropriate bin.

The F10.7 number determines the mean auroral frequency with a high degree of precision (correlation coefficient of -0.96) under sunlit conditions (Fig. 2a). Under dark conditions there is no correlation (-0.01) between the F10.7 number and auroral frequency (Fig. 2b). Taken together, these results are very powerful evidence that it is not some hidden variable indirectly related to solar cycle (such as solar-wind velocity or density) that produces the F10.7 effect but the presence of ionizing solar radiation itself. Indeed, we computed the hourly average values of various solar-wind parameters observed by the NASA IMP-8 satellite from 1984 to 1995; they scarcely varied. The high correlation coefficient under sunlit conditions does not mean that auroral frequency is uniquely determined by the F10.7 number (for example, in a large sample size, the mean height of children in a given age bin will be highly correlated with age, yet age alone does not determine an individual child's height).

No single theory for auroral arc formation has widespread acceptance⁹, although the results are increasingly constraining. The electron acceleration process occurs between approximately 1–3 Earth radii above the surface of the Earth, and occurs in upward regions of field-aligned currents flowing between the ionosphere and the magnetosphere^{10,11}. Most recently, it has been shown that intense aurorae, which occur most often in the 18:00–24:00 MLT sector, are three times less frequent when this sector is under sunlit conditions than in darkness⁴. These findings have previously led us to endorse the ionospheric feedback mechanism for auroral arc formation¹². The dynamics of magnetospheric plasma, including magnetic substorms, require geomagnetic field-aligned currents driven from near-Earth space into, across and back out of the ionosphere. The ionospheric conductivity feedback hypothesis starts by noting that an unstable situation exists when insufficient conductivity exists in the ionosphere to support these imposed currents. The rapidly changing density gradients that exist above the ionosphere can make electromagnetic wave resonances possible. The resulting acceleration of electrons to keV energies creates sufficient ionospheric conductivity to complete the circuit. Moreover, once a localized region of enhanced ionospheric conductivity is started, more current is directed towards that region, carried by additional precipitating electrons, which then further enhance the localized conductivity. The findings reported here strongly support the ionospheric conductivity feedback mechanism for auroral arc formation. □

Received 10 February; accepted 17 March 1998.

1. Schröder, W. *Das Phänomen des Polarlichts* (Wissenschaftliche Buchgesellschaft, Darmstadt, 1984).
2. Chamberlain, J. W. *The Physics of Aurora and Airglow* (American Geophysical Union, Washington DC, reprinted 1996).
3. Barth, C. A. in *The Upper Mesosphere and Lower Thermosphere: A Review of Experiment and Theory* (ed. R. M. Johnson, R. M. & Killeen, T. L.) 225–233 (American Geophysical Union, Washington DC, 1995).
4. Newell, P. T., Meng, C.-I. & Lyons, K. M. Suppression of discrete aurora by sunlight. *Nature* **381**, 766–767 (1996).
5. Newell, P. T., Lyons, K. M. & Meng, C.-I. A large survey of electron acceleration events. *J. Geophys. Res.* **101**, 2599–2614 (1996).
6. Hardy, D. A. *et al.* Report no. AFGL-TR-84-0317 (Air Force Geophysics Laboratory, Boston, MA, 1984).
7. Evans, D. S. The observations of a near monoenergetic flux of auroral electrons. *J. Geophys. Res.* **73**, 2315–2323 (1968).
8. Rasmussen, C. E., Schunk, R. W. & Wickwar, V. B. A photochemical equilibrium model for ionospheric conductivity. *J. Geophys. Res.* **93**, 9831–9840 (1988).
9. Borovsky, J. E. Auroral arc thicknesses as predicted by various theories. *J. Geophys. Res.* **98**, 6101–6138 (1993).
10. Lyons, L. R. Generation of large-scale regions of auroral currents, electric potentials, and precipitation by the divergence of the convection electric field. *J. Geophys. Res.* **85**, 17–24 (1980).
11. Block, L. P. & Falthammer, C.-G. in *Auroral Physics* (eds Meng, C.-I., Rycroft, M. J. & Frank, L. A.) 109–118 (Cambridge Univ. Press, 1991).
12. Lysak, R. J. Feedback instability of the ionospheric resonant cavity. *J. Geophys. Res.* **96**, 1553–1568 (1991).

Acknowledgements. We thank D. Hardy (Phillips Laboratory) and colleagues for designing and building the DMSP particle detectors and for sharing the data. The DMSP data were obtained from NOAA in Boulder, Colorado and from D. Hardy; the F10.7 data were obtained from NOAA. IMP-8 solar-wind data are from Goddard Space Flight Center, courtesy of R. Lepping and A. Lazarus.

Correspondence and requests for materials should be addressed to P.T.N.

Implementation of a quantum search algorithm on a quantum computer

Jonathan A. Jones^{*†}, Michele Mosca^{*‡} & Rasmus H. Hansen[†]

^{*} Oxford Centre for Molecular Sciences, New Chemistry Laboratory, South Parks Road, Oxford OX1 3QT, UK

[†] Centre for Quantum Computation, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

[‡] Mathematical Institute, 24–29 St Giles', Oxford OX1 3LB, UK

In 1982 Feynman¹ observed that quantum-mechanical systems have an information-processing capability much greater than that of corresponding classical systems, and could thus potentially be used to implement a new type of powerful computer. Three years later Deutsch² described a quantum-mechanical Turing machine, showing that quantum computers could indeed be constructed. Since then there has been extensive research in this field, but although the theory is fairly well understood, actually building a quantum computer has proved extremely difficult. Only two methods have been used to demonstrate quantum logic gates: ion traps^{3,4} and nuclear magnetic resonance (NMR)^{5,6}. NMR quantum computers have recently been used to solve a simple

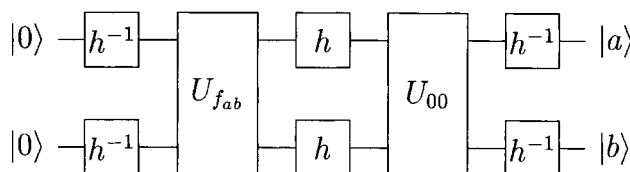


Figure 1 A quantum circuit for the implementation of a quantum search algorithm on a two-qubit computer. Gates marked h (pseudo-Hadamard gates) act to take a single eigenstate to a uniform superposition of the four possible eigenstates, whereas gates marked h^{-1} implement the inverse operation. The first two qubit gate U_{fab} corresponds to an evaluation of the function f_{ab} , replacing an eigenstate $|ij\rangle$ by $-|ij\rangle$ if $i = a$ and $j = b$, whereas U_{00} simply replaces $|00\rangle$ by $-|00\rangle$.

quantum algorithm—the two-bit Deutsch problem^{7,8}. Here we show experimentally that such a computer can be used to implement a non-trivial fast quantum search algorithm initially developed by Grover^{9,10}, which can be conducted faster than a comparable search on a classical computer.

Among other applications, Grover's algorithms enable an extremely rapid search over the domain of a binary function to find elements for which this function is satisfied (that is, the function has the value 1). This approach is simpler if the number of satisfying values is known beforehand, and is particularly simple when precisely one-quarter of the elements in the domain satisfy the function¹¹. The algorithm can be demonstrated by using a computer with two quantum bits (qubits) to search a two-bit domain in which one of the four elements satisfies the function. A classical search of this domain would require between one and three evaluations of the function to find the satisfying element, whereas a quantum search can find this element with only one function evaluation. In the more general case of searching a domain of size N for one of k satisfying elements, the classical search requires about $\frac{1}{2}(N/k)$ function evaluations, whereas the quantum search¹¹ requires only $O(\sqrt{N/k})$.

A quantum circuit for implementing a quantum search in a two-qubit system is shown in Fig. 1. This algorithm uses a single function evaluation to label the single state that satisfies the function, followed by a series of gates that drive the system into this particular state. There are four possible functions f with a unique satisfying element, and these functions are conveniently labelled by the bit pattern of the satisfying element. The algorithm starts with the quantum computer in state $|00\rangle$ and ends with the computer in a state corresponding to the bit pattern of the unique element, and so this element can be immediately identified by determining the final state of each qubit.

This algorithm was implemented with our two-qubit NMR quantum computer, described in ref. 7. This computer uses the two spin states of ^1H nuclei in a magnetic field as qubits, and radio-frequency (RF) fields and spin–spin couplings between the nuclei are used to implement quantum logic gates. Our molecule, partially deuterated cytosine, contains two ^1H nuclei, and thus can be used to implement a two-qubit computer. The pseudo-Hadamard gates (h) were implemented by using 90° pulses; the function evaluation was performed with the pulse sequence shown in Fig. 2. The phases of five of the pulses depend on which of the four possible functions is to be implemented, and these phases should be set as shown in the figure caption. The final gate, U_{00} , is easily implemented, as it is identical to $U_{f_{00}}$.

The algorithm should start with the computer in state $|00\rangle$, but with an NMR quantum computer it is not practicable to begin in a true $|00\rangle$ state. By using the methods of Cory *et al.*⁵ it is, however, possible to create an effective pure state, which behaves in an

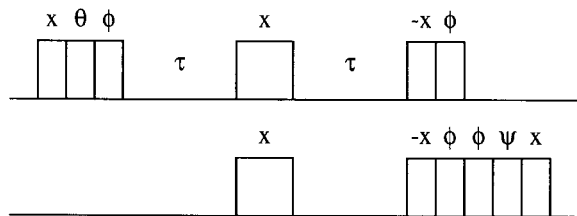


Figure 2 NMR pulse sequence used to implement $U_{f_{ab}}$. Narrow boxes correspond to 90° pulses, whereas wide boxes are 180° pulses; the upper and lower lines refer to the nuclear spins corresponding to the first and second qubits, respectively. The time period τ , during which no pulses are applied, is set equal to $1/4J$, where J is the size of the spin–spin coupling between the nuclei. The phase of each pulse is written above it, and for five of the pulses this phase depends on which of the four functions f_{ab} is to be implemented. These phases should be set as follows: f_{00} , $\theta = +y$, $\phi = +x$, $\psi = -y$; f_{01} , $\theta = +y$, $\phi = -x$, $\psi = +y$; f_{10} , $\theta = -y$, $\phi = -x$, $\psi = -y$; f_{11} , $\theta = -y$, $\phi = +x$, $\psi = +y$.

equivalent manner. Similarly it is not practicable to determine the final state directly, but an equivalent measurement can be made by exciting the spin system with a further 90° pulse and observing the phases of the resulting NMR signals. The absolute phase of an NMR signal depends in a complex manner on a variety of experimental details, so it is not possible to interpret absolute phases, but this can be overcome by measuring a reference signal, obtained by applying a 90° pulse directly to the initial state.

The results of this approach are shown in Fig. 3. Five spectra are shown: a reference spectrum acquired by using a single 90° pulse, and spectra acquired from the same computer implementing the search algorithm for each of the four possible functions, f . Each spectrum consists of two closely spaced pairs of lines: each pair of lines corresponds to a single qubit, whereas the barely visible splitting within each pair arises from the spin–spin coupling, J , used to implement the two-qubit gates. To improve the appearance of the spectra the final 90° detection pulse was preceded by a magnetic field gradient pulse, which acts to dephase most of any error terms that might occur.

The reference spectrum corresponds to the computer's being in state $|00\rangle$, and the phase of this spectrum was adjusted so that both lines are in positive absorption phase (that is, pointing upwards). The same phase correction was then applied to the other four spectra, allowing positive absorption lines to be interpreted as qubits in state $|0\rangle$, whereas negative absorption lines can be interpreted as qubits in state $|1\rangle$. The left-hand pair of lines arises from the first spin, and thus corresponds to the first qubit; the right-hand pair of lines corresponds to the second qubit. Thus, for example, spectrum c in Fig. 3 corresponds to the state $|01\rangle$. Examining the four spectra b–e, it is clear that our implementation of a quantum search with the function f_{ab} leaves the computer largely in a final state $|ab\rangle$, much as expected.

There are, however, small but significant errors in the calculation, which result in distortions in the final spectra. Although most of these distortions are removed by the field gradient pulse, their effects remain visible as variations in the heights of the NMR lines. These errors arise from a variety of causes, but the most significant is errors in the NMR pulse sequence used to implement the calculation.

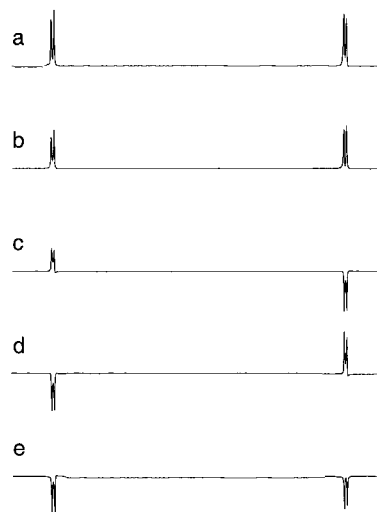


Figure 3 Experimental spectra from our NMR quantum computer. Spectrum a is a reference spectrum, used to determine the absolute phases of the NMR signals; spectra b–e were acquired from the same computer implementing the quantum search algorithm using each of the four possible functions: b, f_{00} ; c, f_{01} ; d, f_{10} ; e, f_{11} . These four spectra were processed using the reference phase obtained from spectrum a, so the phases of the signals can be interpreted as states of the corresponding qubits, as described in the text.

tion. These pulse sequences require a large number of selective pulses, that is, pulses affecting only one of the two spins making up the computer. In practice it is difficult to achieve the desired effect at one spin while leaving the other entirely unaffected, resulting in errors in the final result. Interestingly the distortions are much worse in some cases than in others: they are particularly bad in the spectrum obtained when the function is f_{01} . Understanding this variability could lead to techniques for reducing the distortion in all cases.

Implementing Grover's algorithm is a major step forward for NMR quantum computing (since this Letter was first submitted, an implementation of Grover's algorithm by using heteronuclear NMR has been published¹²), but is by no means the limit of what can be achieved. Preliminary studies have been made of systems containing three qubits¹³, and it should be possible to build larger NMR quantum computers, allowing the implementation of more complex algorithms. □

Received 6 March; accepted 23 April 1998.

1. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
2. Deutsch, D. Quantum-theory, the Church–Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).
3. Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
4. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4715 (1995).
5. Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **94**, 1634–1639 (1997).
6. Gershenfeld, N. A. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
7. Jones, J. A. & Mosca, M. Implementation of a quantum algorithm to solve Deutsch's problem on a nuclear magnetic resonance quantum computer. *J. Chem. Phys.* (in the press); also LANL preprint quant-ph/9801027.
8. Chuang, I. L., Vandersypen, L. M. K., Zhou, X., Leung, D. W. & Lloyd, S. Experimental realization of a quantum algorithm. *Nature* **393**, 143–146 (1998); also LANL preprint quant-ph/9801037.
9. Grover, L. K. in *Proc. 28th Annual ACM Symp. on Theory of Computation* 212–218 (ACM, New York, 1996).
10. Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
11. Boyer, M., Brassard, G., Høyer, P. & Tapp, A. in *Proc. 4th Workshop on Physics and Computation—PhysComp '96* (eds Toffoli, T., Biafore, M. & Leão, J.) 36–41 (New England Complex Systems Institute, Boston, MA, 1996).
12. Chuang, I. L., Gershenfeld, N. & Kubinec, M. Experimental implementation of fast quantum searching. *Phys. Rev. Lett.* **80**, 3408–3411 (1998).
13. Laflamme, R., Knill, E., Zurek, W. H., Catásti, P. & Mariappan, S. V. S. NMR GHZ. *Proc. R. Soc. Lond. A* (in the press).

Acknowledgements. We thank A. Ekert for discussions; J.A.J. thanks C. M. Dobson for encouragement. This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by the UK EPSRC, BBSRC and MRC. M.M. thanks CESC (UK) for their support. R.H.H. thanks the Danish Research Academy for financial assistance.

Correspondence and requests for materials should be addressed to J.A.J. (e-mail: jones@bioch.ox.ac.uk).

Carbon nanotubule membranes for electrochemical energy storage and production

Guangli Che, Brinda B. Lakshmi, Ellen R. Fisher & Charles R. Martin

Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523, USA

Ensembles of aligned and monodisperse tubules of graphitic carbon can be prepared by a templating method^{1–4} that involves the chemical-vapour deposition of carbon within the pores of alumina membranes^{5–7}. Tubules with diameters as small as 20 nm have been prepared in this way^{7,8}. The carbon comprising these tubules can be transformed from a disordered material to very highly ordered graphite⁹. Here we show that template-synthesized

carbon tubules can be fabricated as free-standing nanoporous carbon membranes, and that narrower, highly ordered graphitic carbon nanotubes can be prepared within the membrane's tubules. Both the outer and the inner tubules are electrochemically active for intercalation of lithium ions, suggesting possible applications in lithium-ion batteries^{9,10}. The membranes can also be filled with nanoparticles of electrocatalytic metals and alloys. Such catalyst-loaded membranes can be used to electrocatalyse O₂ reduction and methanol oxidation, two reactions of importance to fuel-cell technology.

Carbon nanostructures are of tremendous interest^{10,11}, from both a fundamental and an applied perspective. Applications investigated include use for storage of hydrogen¹² and other gases¹³, as a catalyst support^{10,11,14,15} and as a tip for scanning probe microscopy^{16,17}. In addition, participants at the recent Rice University Workshop on these structures have suggested¹¹ that these tubes might be used as membrane materials for batteries and fuel cells, anodes for lithium-ion batteries, capacitors, and chemical filters. Many of these proposed applications will require alignment^{11,18} of the nanotubes and aggregation of these aligned tubes to form a membrane. We have accomplished these objectives and have used the resulting aligned carbon-tubule membranes to demonstrate some of these proposed energy-related applications.

The previously described chemical-vapour deposition (CVD) method⁵ was used to synthesize the carbon tubules within the pores (200-nm-diameter) of a commercially available alumina template membrane^{1,5}. (Although this method has been used to make carbon tubules with diameters as small as 20 nm (ref. 7; G.C. and C.R.M., unpublished results), we have found it a good strategy to begin with membranes with larger pore diameters.) This membrane has 60% porosity and is ~60 µm thick. In addition to producing a carbon tubule within each pore, the CVD method yields carbon surface films (~20 nm thick) that cover both faces of the alumina membrane⁵. This is important because these surface films will be used to hold the tubules together, in an aligned fashion, after removal of the underlying alumina membrane (see below).

For all of the energy-related applications described here, catalytic metal nanoparticles were then prepared within the CVD template-synthesized carbon tubules. This was accomplished by immersion of the C/alumina membrane into a solution of the desired metal ion, either 73 mM H₂PtCl₆ (aqueous)⁷, 37 mM H₂PtCl₆ plus 73 mM RuCl₃ (aqueous), or 124 mM Fe(NO₃)₃ (ethanolic)¹⁸. After immersion, the membrane was dried in air, and the ions were reduced to the corresponding metal or alloy by a 3-h exposure to flowing H₂ gas at 580 °C. The underlying alumina was then dissolved away in 46% HF solution to obtain the desired free-standing carbon-tubule membrane. The alumina dissolution process can be followed visually because the high temperature (900 °C) used in the CVD synthesis⁵ causes the alumina to curl into a tight cylinder. As the underlying alumina is dissolved by the HF, this cylinder relaxes, ultimately producing the desired planar, free-standing, carbon-tubule membrane, suspended in the HF solution. This membrane was removed from the HF solution and rinsed with water.

The scanning electron micrograph in Fig. 1a shows that the carbon tubules that make up these membranes are aligned, and that the ends of these tubules are open. Fig. 1b shows a transmission electron micrograph of a single tubule that was removed (by ultrasonication) from such a membrane. The tubule walls are sufficiently thin that the Pt/Ru nanoparticles contained within can be seen. Images of this type were used to obtain the following particle size distributions: Pt (7.1 ± 0.4 nm), Pt/Ru (1.59 ± 0.03 nm). Iron nanoparticles of this type can be used as catalysts for the CVD growth of highly ordered graphitic-carbon nanotubes^{5,18}. This approach was used to grow such nanotubes within the template-synthesized tubules. Figure 1c shows that graphitic nanotubes that wind their way through the template-synthesized tubules are obtained.

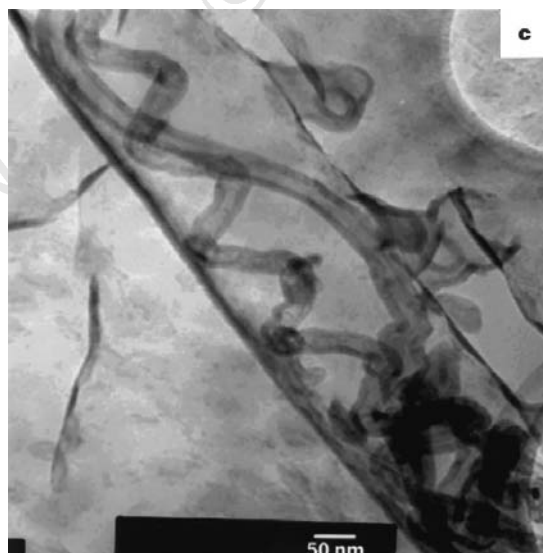
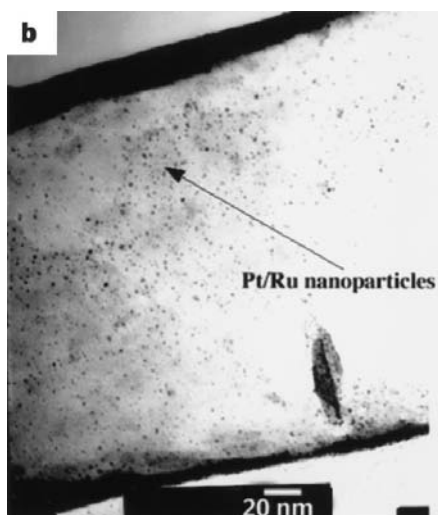
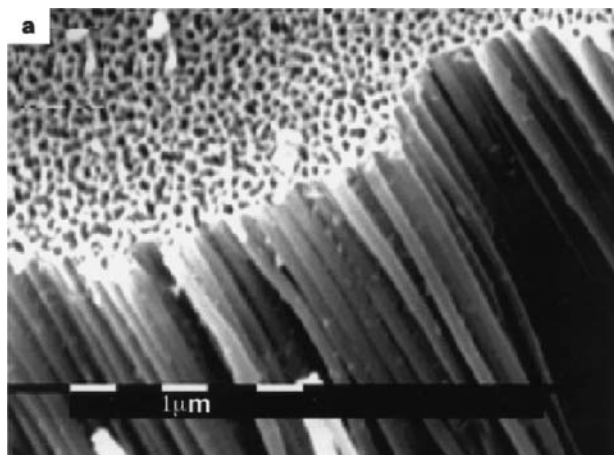


Figure 1 Electron micrographs of tubules and membranes. **a**, Scanning electron micrograph of a template-synthesized carbon tubule membrane after dissolving the alumina template. The edge (lower part of the image) showing the tubules and surface (upper part of the image) are shown. The 20-nm-thick surface film cannot be seen at this magnification. **b**, Transmission electron micrograph of a Pt/Ru nanoparticle-filled carbon tubule. **c**, Transmission electron micrograph of the Fe-catalysed CVD graphitic carbon inner nanotubes within the outer template-synthesized tubules (tube-in-tube membrane).

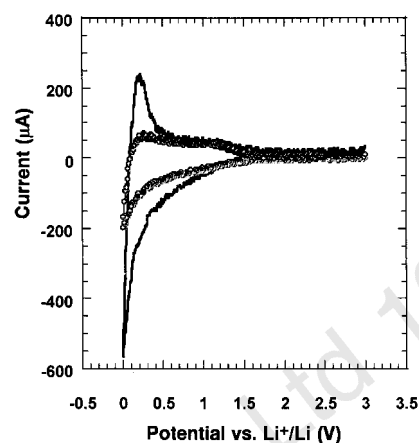


Figure 2 Cyclic voltammograms illustrating possible Li-ion battery application of the carbon nanotubule membranes. Cyclic voltammograms associated with the reversible intercalation of Li^+ into a tube-in-tube membrane (solid curve) and into the template-synthesized tubule membrane (plotting symbols). Scan rate, 0.3 mV s^{-1} .

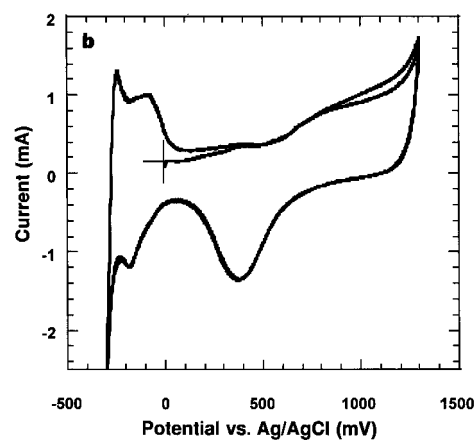
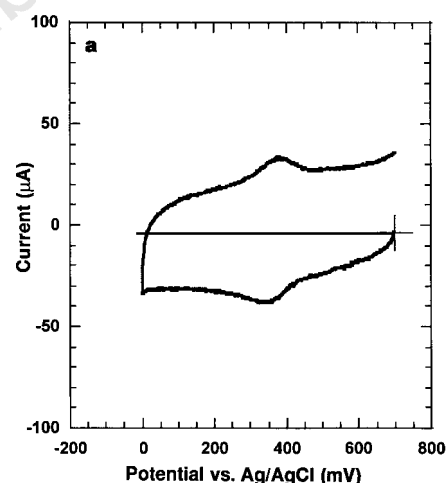


Figure 3 Cyclic voltammograms (CVs) of template-synthesized carbon-tubule membranes. Scan rate, 50 mV s^{-1} . **a**, The outer CV is for the nanotubule membrane before deposition of the Pt nanoparticles into the tubules. The inner CV is for a glassy carbon electrode of equivalent geometric area ($0.05 \text{ M H}_2\text{SO}_4$). **b**, Nanotubule membranes after deposition of the Pt nanoparticles ($1.0 \text{ M H}_2\text{SO}_4$).

We have been exploring the possibility of using high-surface-area tubular nanostructures of Li^+ insertion materials as electrodes in Li-ion batteries^{19,20}. Because the anode material of choice for such batteries is carbon, we were interested in learning whether the carbon-tubule membranes can reversibly intercalate Li^+ , and if the Fe-catalysed nanotubes contained within (Fig. 1c) add an extra increment of Li^+ intercalation capacity. The desired membrane was attached to a copper wire with silver epoxy and this adhesive was covered with inert Torr-Seal epoxy. (The underlying alumina membrane was not removed for these, and only these, studies.) The membrane was then immersed in a solution that was 1 M in LiClO_4 in 30:70 (v/v) ethylene carbonate/diethyl carbonate along with a Li-foil counter and reference electrode. Reversible Li^+ intercalation was studied using cyclic voltammetry and constant-current discharge experiments^{20–22}.

The cyclic voltammograms (CVs) in Fig. 2 show that the carbon-tubule membranes can, indeed, reversibly intercalate Li^+ and that the Fe-catalysed nanotubes contained within (tube-in-tube membrane, Fig. 1c) more than doubles the intercalation capacity. The inner (lower-current) CV shows the reversible intercalation of Li^+ into the carbon-tubule membrane before synthesis of the inner nanotubes; this CV is characteristic of Li^+ -intercalation into disordered C (ref. 21). The outer CV is for the analogous tube-in-tube membrane (Fig. 1c). The higher current shows that the capacity for the tube-in-tube membrane is, indeed, higher; and the pronounced peak in the return wave indicates that this extra capacity is associated with highly ordered carbon^{8,22}. The template-

synthesized tubule membrane shows a high Li^+ intercalation capacity (490 mA h g^{-1}) and the data in Fig. 2 indicate that, on a volumetric basis, the capacity of the tube-in-tube membrane is at least two times higher. This is, to our knowledge, the first time that reversible Li^+ intercalation has been demonstrated for such carbon nanotubules.

We also have a long-standing interest in the electrocatalysis of fuel-cell reactions²³, and it seemed likely that the Pt and Pt/Ru-loaded carbon-tubule membranes might be excellent electrocatalysts for such reactions. To explore this point, a Pt or Pt/Ru-loaded membrane (after removal of the underlying Al_2O_3) was placed onto the surface of a glassy carbon electrode (area 0.07 cm^2 ; Bioanalytical Systems, W. Lafayette, Indiana) and one drop of a 5% Nafion solution (Aldrich) was applied. The solvent was allowed to evaporate, and the Nafion acted as an adhesive to hold the membrane to the electrode surface.

We look first at the electrocatalysis of O_2 reduction using the Pt nanoparticle-containing membranes. Figure 3 shows CVs, in the absence of O_2 , for carbon-tubule membranes before and after deposition of the Pt nanoparticles within the tubules. Before deposition, only background currents, characteristic of carbon electrodes, are seen (Fig. 3a). We note that the background currents at the nanotube membrane electrode (outer CV) are much greater than at a glassy carbon electrode (inner CV that appears as a flat line) of equivalent geometric area. This is because the inner surfaces of the carbon tubules are electrochemically accessible. This greater electrochemically active surface area is a great advantage for any electrocatalytic application.

After Pt deposition, clear and characteristic Pt surface electrochemistry²⁴ is observed (Fig. 3b). Figure 4a compares O_2 reduction currents for the empty and Pt-nanoparticle-loaded carbon-tubule membranes. In the absence of the Pt nanoparticles, no O_2 reduction is observed over this potential window (curve B). After deposition of the nanoparticles, a very large O_2 reduction wave, at potentials characteristic for Pt electrocatalysis in this solution²⁵, is observed (curve A). As would be expected from the higher electrochemically active surface area, these currents are 20 times higher than currents observed at a Nafion-coated Pt electrode of equivalent geometric area. Figure 4b shows analogous results for the electrocatalysis of methanol oxidation. Before deposition of the Pt/Ru nanoparticles no methanol oxidation is observed. After deposition very large and characteristic²⁶ methanol oxidation waves are seen.

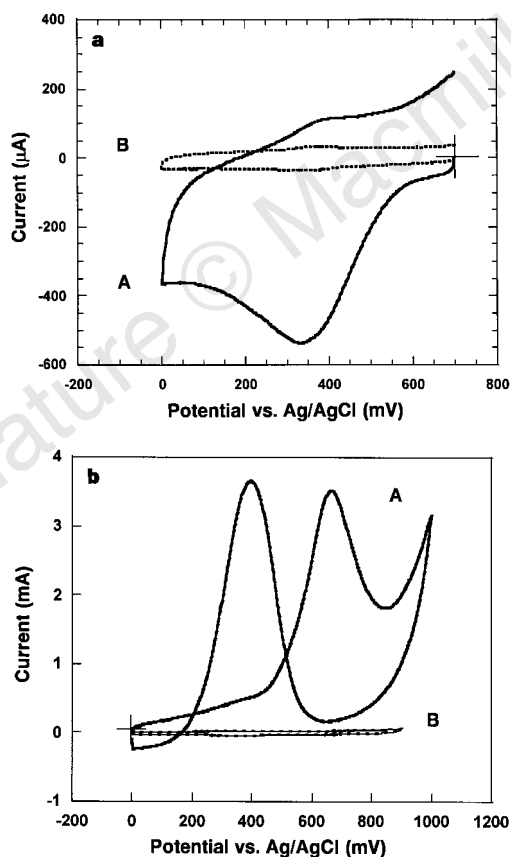


Figure 4 Cyclic voltammograms illustrating electrocatalytic applications of the carbon nanotubule membranes. **a**, Cyclic voltammograms for O_2 reduction at template-synthesized carbon-tubule membranes in $0.05 \text{ M H}_2\text{SO}_4$. The solution was O_2 saturated. Trace A, after deposition of the Pt nanoparticles. Trace B, before deposition of the Pt nanoparticles. **b**, Cyclic voltammograms for methanol oxidation (2 M methanol in 1 M aqueous H_2SO_4). Trace A, after deposition of the Pt/Ru nanoparticles. Trace B, before deposition of the Pt/Ru nanoparticles. Scan rate, 50 mV s^{-1} . Geometric area of nanotubule membranes, 0.07 cm^2 .

Methods

The previously described CVD method⁵ (ethylene precursor) was used to synthesize the carbon tubule membranes. No metal catalyst was used, and the tubules were not subsequently graphitized⁵. The highly ordered graphitic carbon tubules were synthesized using the method described elsewhere^{5,18}. Li -intercalation capacity was determined using both CV (Fig. 2) and constant-current discharge experiments (potential limits, 1.5 and 0.0 V). The CV-determined capacity was constant over a scan rate region from 0.1 to 1 mV s^{-1} . The mass of carbon used was determined gravimetrically. As is typically the case for C anodes, irreversible capacity was observed on first intercalation; the capacity quoted is from subsequent charge/discharge experiments. Capacity did not fade over 20 charge/discharge cycles.

Received 20 February; accepted 20 April 1998.

- Martin, C. R. Nanomaterials—A membrane-based synthetic approach. *Science* **266**, 1961–1966 (1994).
- Nishizawa, M., Menon, V. P. & Martin, C. R. Metal nanotube membranes with electrochemically switchable ion-transport selectivity. *Science* **268**, 700–702 (1995).
- Parthasarathy, R. V. & Martin, C. R. Synthesis of polymeric microcapsule arrays and their use for enzyme immobilization. *Nature* **369**, 298–301 (1994).
- Jirage, K. B., Hulteen, J. C. & Martin, C. R. Nanotube-based molecular filtration membranes. *Science* **278**, 655–658 (1997).
- Che, G., Lakshmi, B. B., Martin, C. R., Fisher, E. R. & Ruoff, R. S. Chemical vapor deposition based synthesis of carbon nanotubes and nanofibers using a template method. *Chem. Mater.* **10**, 260–267 (1998).
- Kyotani, T., Tsai, L.-F. & Tomita, A. Formation of ultrafine carbon tubes by using an anodic aluminum oxide film as a template. *Chem. Mater.* **7**, 1427–1428 (1995).

7. Kyotani, T., Tsai, L.-F. & Tomita, A. Formation of platinum nanorods and nanoparticles in uniform carbon nanotubes prepared by a template carbonization method. *Chem. Commun.* 701–702 (1997).
8. Dahu, J. R., Zheng, T., Liu, Y. & Xue, J. S. Mechanisms for lithium insertion in carbonaceous materials. *Science* **270**, 590–593 (1995).
9. Sato, K., Noguchi, M., Demachi, A., Oki, N. & Endo, M. A mechanism of lithium storage in disorder carbons. *Science* **264**, 556–558 (1994).
10. Freemantle, M. Filled carbon nanotubes could lead to improved catalysts and biosensors. *Chem. Eng. News* **74**, 62 (1996).
11. Baum, R. M. Nurturing nanotubes, remarkable properties of nanotubes suggest range of new technologies; workshop examines how to develop them. *Chem. Eng. News* **75**, 39–41 (1997).
12. Dillon, A. C. *et al.* Storage of hydrogen in single-walled carbon nanotubes. *Nature* **386**, 377–379 (1997).
13. Gadd, G. E. *et al.* The world's smallest gas cylinders. *Science* **277**, 933–936 (1997).
14. Planeix, J. M. *et al.* Application of carbon nanotubes as supports in heterogeneous catalysis. *J. Am. Chem. Soc.* **116**, 7935–7936 (1994).
15. Sloan, J., Cook, J., Green, M. L. H., Hutchison, J. L. & Tenne, R. Crystallisation inside fullerene related structures. *J. Mater. Chem.* **7**(7), 1089–1095 (1997).
16. Wong, S. S., Harper, J. D., Lansbery, P. T. & Lieber, C. M. Carbon nanotube tips: High-resolution probes for imaging biological systems. *J. Am. Chem. Soc.* **120**, 603–604 (1998).
17. Dai, H. J., Hafner, J. H., Rinzler, A. G., Colbert, D. T. & Smalley, R. E. Nanotubes as nanopores in scanning probe microscopy. *Nature* **384**, 147–150 (1996).
18. Li, W. Z. *et al.* Large-scale synthesis of aligned carbon nanotubes. *Science* **274**, 1701–1703 (1996).
19. Nishizawa, M., Mukai, K., Kuwabata, S., Martin, C. R. & Yoneyama, H. Template synthesis of polypyrrole-coated spinnable LiMn_2O_4 nanotubes and their properties as cathode active materials for lithium batteries. *J. Electrochem. Soc.* **144**, 1923–1927 (1997).
20. Che, G., Jirage, K. B., Fisher, E. R., Martin, C. R. & Yoneyama, H. Chemical-vapor deposition-based template synthesis of microtubular TiS_2 battery electrodes. *J. Electrochem. Soc.* **144**, 4296–4302 (1997).
21. Jean, M., Desnoyer, C., Tranchant, A. & Messina, R. Electrochemical and structural studies of petroleum coke in carbonate-based electrolytes. *J. Electrochem. Soc.* **142**, 2122–2125 (1995).
22. Verbrugge, M. W. & Koch, B. J. Lithium intercalation of carbon-fiber microelectrodes. *J. Electrochem. Soc.* **143**, 24–31 (1996).
23. Bockris, J. O'M. & Srinivasan, S. *Fuel Cells: their Electrochemistry* (McGraw-Hill, New York, 1969).
24. Bard, A. J. & Faulkner, L. R. *Electrochemical Methods, Fundamentals and Applications* (Wiley, New York, 1980).
25. Ye, S., Ujih, A. K. & Dao, L. H. A new fuel cell electrocatalyst based on highly porous carbonized polyacrylonitrile foam with low platinum loading. *J. Electrochem. Soc.* **143**, L7–L9 (1996).
26. Shen, P. K. & Tseung, A. C. C. Anodic oxidation of methanol on Pt/WO_3 in acidic media. *J. Electrochem. Soc.* **141**, 3082–3089 (1994).

Acknowledgements. This work was supported by the US Department of Energy and the US Office of Naval Research.

Correspondence and requests for materials should be addressed to C.R.M. (e-mail: crmartin@lamar.colostate.edu).

Entropically driven microphase transitions in mixtures of colloidal rods and spheres

Marie Adams*, Zvonimir Dogic*, Sarah L. Keller† & Seth Fraden*

* The Complex Fluids Group, Martin Fisher School of Physics, Brandeis University, Waltham, Massachusetts 02254, USA

† Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA

Although the idea that entropy alone is sufficient to produce an ordered state is an old one in colloid science¹, the notion remains counter-intuitive and it is often assumed that attractive interactions are necessary to generate phases with long-range order. The phase behaviour for both rods and spheres has been studied experimentally^{1–7}, theoretically^{8,9} and by computer simulations¹⁰. Here we describe the phase behaviour of mixtures of colloidal rod-like and sphere-like particles (respectively viruses and polystyrene latex or polyethylene oxide polymer) under conditions in which they act like hard³ particles^{2,3}. We find a wealth of behaviour: bulk demixing into rod-rich and rod-poor phases and microphase separation into a variety of morphologies. One microphase consists of layers of rods alternating with layers of spheres¹¹; in another microphase of unanticipated complexity, the spheres reversibly assemble into columns, which in turn pack into a crystalline array. Our experiments, and previous theory and computer simulations¹¹, suggest that this phase behaviour is

entropically driven by steric repulsion between particles. The phenomena are likely to be quite general, applying also for example to low-molecular-mass liquid crystals¹². This kind of microphase separation might also be relevant to systems of amphiphiles¹³ and block copolymers¹⁴, to bioseparation methods and DNA partitioning in prokaryotes¹⁵, and to protein crystallization^{16,17} and the manufacture of composite materials.

The phase behaviour of a colloidal suspension is found by minimizing the free energy $F = U - TS$. Hard particles are forbidden to interpenetrate and the interaction energy U is zero as long as the particles are not in contact. Thus phase behaviour is determined by maximizing the entropy S . For hard spheres, imagine a suspension of billiard balls, and for hard rods substitute billiard cues. Examples of entropically driven ordering exhibited by hard rods are the nematic and smectic liquid-crystalline phases, whereas for hard spheres a fluid–crystal transition occurs. These phase transitions have been extensively studied theoretically^{8,9}, by computer simulations¹⁰, and experimentally in a variety of colloidal systems, including the ones studied here^{1–7}. Another class of entropically driven ordering arises in mixtures in which demixing transitions often occur. If one of the physical properties of the individual molecular species, such as shape or flexibility, is different enough, then either bulk or microphase separation will occur for an appropriate composition of the components.

The physical origin of phase separation in hard particle suspensions is both extremely simple and general. The free volume of a suspension composed of a binary mixture of macromolecules is maximized when the two components phase-separate. Increasing the free volume (see insert to Fig. 1) raises the translational entropy of the macromolecules, but at the expense of lowering the entropy of mixing. At low concentrations, the mixing entropy dominates and the solution is miscible, but with increasing concentration, if the gain in free volume is sufficient, phase separation will occur. Although the interparticle potential is purely repulsive, like-components experience an effective attraction. This phenomenon is known by different names—depletion attraction in physics and

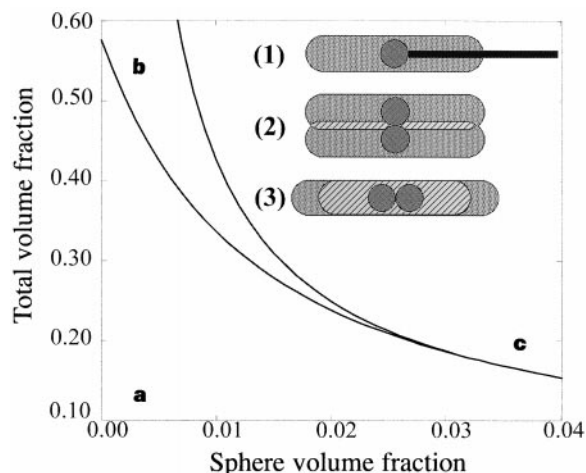


Figure 1 Theoretical phase diagram for aligned spherocylinders of length L , diameter D , and spheres of diameter σ (ref. 11) for $L/D = 100$ and $\sigma/D = 10$. Upon addition of spheres, the total volume fraction needed to form a lamellar phase (b) decreases, hence spheres stabilize the smectic phase. In a, the rods and spheres are miscible, and in c, bulk phase separation occurs. Insert: (1) The volume excluded (light grey) to the centre of a horizontally aligned rod (black) by a sphere (dark grey). (2) Bringing two spheres together in a suspension of parallel rods leads to an overlap of excluded volume (wide cross-hatch)—increasing the free volume and thus the translational entropy of the rod. (3) Aggregation of the spheres parallel to the rods produces more overlap of excluded volume than in insert 2.

chemistry^{16,18–24}, but is macromolecular crowding in biology^{15,25}. Depletion phenomena are general because all particles at any scale possess a hard core that must be considered in any accurate theoretical model. For the case of the colloidal suspensions studied here, hard-core repulsion is the dominant interaction because salt added to the suspension screens electrostatic repulsion.

In our experiments, we either add polyethylene oxide (PEO), polyethylene glycol (PEG) or polystyrene latex (PS) spheres to suspensions of the filamentous bacteriophage fd virus in the isotropic or nematic phases. When low concentrations of spheres that are small compared to the fd length (22 nm diameter PS, 5% volume fraction) are added to a pure fd nematic (20 mg ml⁻¹, 10 mM Tris), we observe bulk phase separation of the rods into an isotropic and nematic phase. The spheres partition into the rod-poor phase in the form of ellipsoidal shaped droplets, with the long axis parallel to the nematic director (rod alignment axis). In contrast, when PS spheres at any concentration larger than about 700 nM are added to the pure fd nematic, complete demixing of the spheres into dense, irregular shaped aggregates elongated along the nematic director is observed.

The full complexity of the phase diagram of rod/sphere mixtures is illustrated in Figs 2 and 3. Upon increasing from zero the concentration of 100-nm-diameter PS spheres in a dilute fd nematic sample, the spheres assemble into columns of ~300 nm diameter that are up to 5 µm long and oriented perpendicular to the rods' long axis. The columns in turn pack on a two-dimensional lattice with diametrically opposed columns approximately two rod-lengths apart. Birefringence measurements indicate no disruption of the alignment of the rods; we call this phase 'columnar'. As the sphere concentration is increased, the columns continuously diminish in diameter and spheres fill in the line joining the two opposite columns that are perpendicular to the rod alignment direction. Eventually the columns disappear and a 'lamellar' phase is formed of a layer of rods alternating with a layer of spheres. Both the lamellar and columnar phases are equilibrium phases; we have one lamellar phase of over two years old.

Freeze-fracture electron microscopy of the lamellar phase (Fig. 3b(2)) revealed cases in which one layer of aligned but disordered rods alternated with a layer of disordered spheres about two diameters thick. This was consistent with optical microscopy (Fig. 3b(1)) results that measured the periodicity of the lamellar phase at 1.1 µm, about 0.2 µm longer than the fd molecule, or twice the diameter of the PS spheres. At higher fd concentrations,

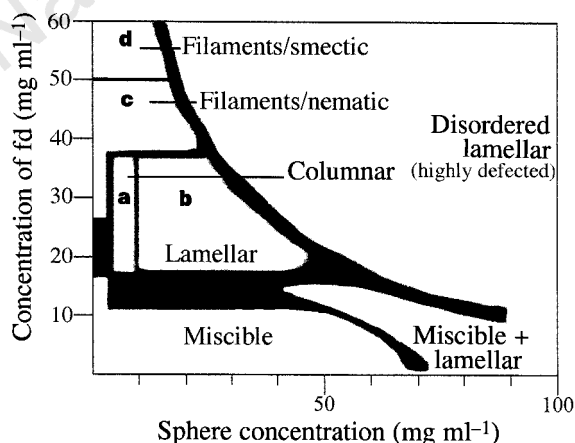


Figure 2 Phase diagram for fd and 100-nm diameter polystyrene spheres. The ionic strength (*I*) was 5 mM at pH 8.2, but a topologically identical diagram was observed with *I* = 50 mM. The phase sequence for pure fd (*I* = 5 mM) is isotropic–nematic at 15 mg ml⁻¹ and nematic–smectic at 55 mg ml⁻¹ (ref. 6). At volume fractions of spheres as low as 0.5% (5 mg ml⁻¹) and rod concentrations of 20 mg ml⁻¹, far below the smectic phase, we observe microphase separation. The appearance and structure of those phases labelled with letters are shown in Fig. 3.

filaments of the lamellar phase were observed. Filaments had uniform diameter, but with a variance of diameter among filaments in the same sample, ranging from 2 to 100 µm diameter with lengths of the order of centimetres. Filaments formed within minutes after initially mixing the fd and PS. Samples ranging in age from a day to several months resembled those shown in Fig. 3c, d, but bulk demixing of the rods and spheres occurred after one year. These filaments formed in co-existence with either fd in a nematic phase, or in the smectic phase. The smectic phase had a periodicity of 0.9 µm, identical to that of pure fd suspensions.

The lamellar phase is initially largely defect-free, and at low concentrations is very flexible. Fluorescently labelled spheres are seen to hop between layers. With increasing sphere concentration, the lamellae become rigid and hopping between layers stops. As the sphere concentration is further increased, the number of defects multiply until the size of defect-free regions is only several molecular lengths. Neither swelling of the layers nor bulk phase separation was noted as a function of sphere concentration.

Phase diagrams of identical topology were observed using polystyrene spheres ranging from 22 to 120 nm diameter, but the concentration of spheres needed to induce lamellar order increased

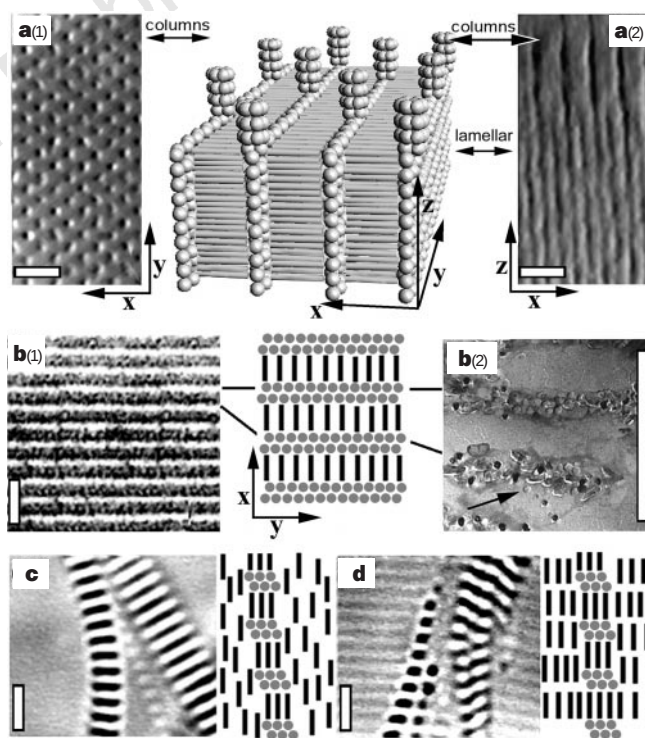


Figure 3 Photographs and diagrams of regions of the phase diagram shown in Fig. 2. Unless specified, photographs are differential interference contrast (DIC) microscopy. The micrographs cannot clearly resolve single particles, so the drawings give only an approximate representation of the structure. The actual volume fraction of the spheres within the columns and lamellae is 10–20%. Length bars are 3 µm. **a(1)**, **a(2)**, Orthogonal sections of co-existing columnar and lamellar phases. **a(1)**, Section in the *x*-*y* plane perpendicular to the columns, corresponding to the upper part of the adjacent drawing—each column is about 3 spheres in diameter; **a(2)**, section in the *x*-*z* plane corresponding to the front of the drawing. The columns (top) have twice the spacing of the lamellae (bottom). The lamellar phase is shown in **b(1)**; **b(2)** is a freeze-fracture electron micrograph²⁸. The arrow indicates a hemispherical divot where a sphere was lifted out of the plane by the fracturing process. In **c** and **d**, filaments of the lamellar phase (high contrast) co-existing with either the phase separated rods in a nematic of uniform texture (**c**) or smectic with weak contrast (**d**). Filaments are metastable but long-lived structures. The periodicity of the lamellar (1.1 µm) is larger than the periodicity of the smectic (0.9 µm).

with decreasing sphere size. Columnar phases were observed for PS in the 60–120 nm range: notably, the diameter of the columns of spheres remained the same at ~ 300 nm. The lamellar phase (but not columnar) was also observed with PEO relative molecular mass (M_r) of 8,000K, which has a hydrodynamic radius of 100 nm (ref. 26).

PS spheres of size greater than 250 nm did not form the lamellar phase. When low concentrations of 300-nm-diameter PS spheres were added to fd nematics, we observed rapid association of the spheres into chain-like structures (Fig. 4). At higher densities, the spheres packed into a cubic array with a lattice constant of one rod length.

A microscopic but highly simplified theory models the fd as perfectly aligned hard rods, the PS as hard spheres, and treats interactions at the second-virial level¹¹. We have used this theory to calculate the stability limits of the phase diagram (Fig. 1). This predicts topologically identical phase diagrams for all sphere diameters. The experimental phase diagrams were remarkably similar to theory for spheres ranging in diameter from 22 to 200 nm. However, above this sphere diameter no lamellar phase is seen, in contrast to theory. The theory also predicts swelling of the layers, in contrast to experiment, and furthermore, the theory does not account for the columnar phase.

The observed lamellar and columnar phases in fd/PS mixtures bear strong resemblance to micellar¹³ and block-copolymer¹⁴ systems, in which microphase ordering is driven by the tendency of the antagonistic (or amphiphilic) portions of an individual molecule to phase separate while being prevented from doing so by chemical bonds. In contrast, no bonds exist between the rods (fd) and spheres (PS), so naively we expect bulk demixing of rods and spheres^{23,24}. The photograph in Fig. 4 of 300-nm spheres immersed in a nematic suspension of fd rods suggests an explanation of why the lamellar phase is preferred to bulk demixing of the rods and spheres^{23,24}. The photographs indicate 300-nm spheres have two preferred configurations: either spheres in contact or separated by one rod length, which suggests that these are low-energy configurations. In other words, the pictures demonstrate that in a rod–sphere mixture there is an effective sphere–sphere attraction (promoting bulk phase separation) and an effective rod–sphere attraction (promoting microphase formation). Depletion theory explains the sphere–sphere association as a way to maximize free volume, and hence total entropy. We speculate that the rod–sphere attraction arises from correlations among the rods induced by the presence of spheres (Fig. 4). These experiments and theoretical arguments demonstrate that the bond linking the two portions of a co-polymer is not essential for inducing microphase ordering, contrary to conventional wisdom^{13,14}.

Although there is partial understanding of the lamellar phase, the

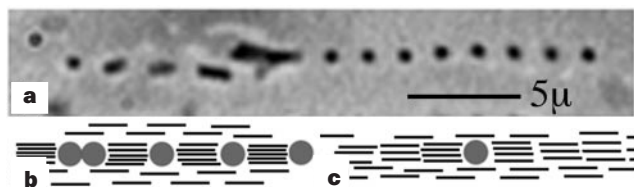


Figure 4 Effective rod–sphere attraction. **a**, DIC micrograph of 300-nm PS spheres in a nematic fd sample. **b**, The spheres either aggregate in close-packed chains (Fig. 1, insert 3) or into open chains in which each sphere is separated by a rod length (0.9 μm). **c**, An interface perpendicular to the nematic alignment direction forces the ends of the rods into registry, setting off a smectic density wave that decays away from the interface. Consequently, slightly more than one rod length away from the sphere the density of rods is lower than average and thus it is easier to insert a second sphere. Such spheres strongly stabilize the lamellar phase (Fig. 1).

columnar phase remains a mystery and presents a theoretical challenge. Analysis of the phenomena described here is relevant to both the high-molecular-mass macromolecules used in this study, as well as low-molecular-mass ones for which interparticle attractions are significant, because all molecules possess a repulsive steric core. Thus the entropic forces discussed here need to be considered in order to have a complete description of the phase behaviour of any mixture containing anisotropically shaped molecules. Furthermore, we have shown that effective attractive potentials of surprising complexity can be constructed through the choice of the shape, size and number of molecules in a suspension, indicating the possibility of manipulating these variables^{16,22} to ‘engineer with entropy’, building order from disorder²¹. □

Methods

The colloidal rods consist of filamentous bacteriophage fd virus, a semiflexible polymer of 6.6 nm diameter, 880 nm contour length, 2.2 μm persistence length, and negative charge density of $10e^-$ per nm (refs 2, 27). Growth and purification of the virus were done using standard methods². The spherical colloids consisted of either monodisperse polystyrene latex spheres, solid charged spheres ranging in diameter from 22 nm to 1.0 μm (Bangs Labs, IDC), or one of either of the water-soluble neutral polymers; polyethylene glycol or polyethylene oxide of relative molecular masses ranging from 8K to 8,000K (Aldrich, Fluka, Sigma). Colloids were suspended in either 10 mM or 100 mM Tris buffer at pH 8.2 (ionic strength, 5 mM or 50 mM, respectively). Samples were flame-sealed in rectangular glass capillaries of 50 μm path length (Vitro Dynamics).

Phase diagrams were determined by direct observation with differential interference contrast (DIC) and polarizing microscopy on a Nikon MicroPhot SA connected to a MTI CCD-72 video camera and a Scion AG-5 video Frame Grabber interfaced with a Macintosh computer running NIH Image.

Received 20 January; accepted 30 March 1998.

- Forsyth, P. A. Jr, Marcelja, S., Mitchell, D. J. & Ninham, B. W. Ordering in colloidal systems. *Adv. Coll. Int. Sci.* **9**, 37–60 (1978).
- Fraden, S. in *Observation, Prediction, and Simulation of Phase Transitions in Complex Fluids* (eds Baus, M., Rull, L. F. & Ryckaert, J. P.) 113–164 (Kluwer Academic, Dordrecht, 1995).
- Poon, W. C. K. & Pusey, P. N. in *Observation—Prediction, and Simulation of Phase Transitions in Complex Fluids* (eds Baus, M., Rull, L. F. & Ryckaert, J. P.) 3–51 (Kluwer Academic, Dordrecht, 1995).
- Tang, J. & Fraden, S. Isotropic–cholesteric phase transition in colloidal suspensions of filamentous bacteriophage fd. *Liq. Cryst.* **19**, 459–467 (1995).
- Tang, J. & Fraden, S. Temperature dependence of the flexibility of fd: Behavior of the isotropic–cholesteric phase boundary and magnetic birefringence. *Biopolymers* **39**, 13–22 (1996).
- Dogic, Z. & Fraden, S. Smectic phase in a colloidal suspension of semiflexible virus particles. *Phys. Rev. Lett.* **78**, 2417–2420 (1997).
- Sato, T. & Teramoto, A. Concentrated solutions of liquid-crystalline polymers. *Adv. Polym. Sci.* **126**, 85–161 (1996).
- Onsager, L. The effects of shape on the interaction of colloidal particles. *Ann. NY Acad. Sci.* **51**, 627–659 (1949).
- Hosino, M., Nakano, H. & Kimura, H. Nematic–smectic transition in an aligned rod system. *J. Phys. Soc. Jpn* **46**, 1709–1715 (1979).
- Frenkel, D. in *Liquids, Freezing, and Glass Transition* (eds Hansen, J. P., Levesque, D. & Zinn-Justin, J.) Ch. 9, 689–762 (Elsevier Science, Amsterdam, 1991).
- Koda, T., Numajiri, M. & Ikeda, S. Smectic–A phase of bidisperse system of parallel hard rods and hard spheres. *J. Phys. Soc. Jpn* **65**, 3551–3556 (1996).
- Rieker, T. P. Organic lyotropic lamellar liquid crystals. *Liq. Cryst.* **19**, 497–500 (1995).
- Gelbart, W. M., Ben-Shaul, A. & Roux, D. (eds) *Micelles, Membranes, Microemulsions, and Monolayers* (Springer, New York, 1994).
- Bates, F. S. Polymer–polymer phase behavior. *Science* **251**, 898–905 (1991).
- Walter, H. & Brooks, D. E. Phase separation in cytoplasm, due to macromolecular crowding, is the basis for microcompartmentation. *FEBS Lett.* **361**, 135–139 (1995).
- ten Wolde, P. R. & Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **277**, 1975–1978 (1997).
- Adams, M. & Fraden, S. Phase behavior of mixtures of rods (tobacco mosaic virus) and spheres (polyethylene oxide, bovine serum albumin). *Biophys. J.* **74**, 669–677 (1998).
- Asakura, S. & Oosawa, F. Interaction between particles suspended in solutions of macromolecules. *J. Polym. Sci.* **33**, 183–192 (1958).
- Flory, P. J. Statistical thermodynamics of mixtures of rodlike particles. 5. Mixtures with random coils. *Macromolecules* **11**, 1138–1141 (1978).
- Gast, A. P., Russel, W. B. & Hall, C. K. An experimental and theoretical study of phase transitions in the polystyrene latex and hydroxyethylcellulose system. *J. Coll. Int. Sci.* **109**, 161–171 (1986).
- Frenkel, D. in *Proceedings of the Sitges Conference on Complex Liquids* (ed. Garrido, L.) 137–148 (Springer, New York, 1993).
- Dinsmore, A. D., Yodh, A. G. & Pine, D. J. Entropic control of particle motion using passive surface microstructures. *Nature* **383**, 239–242 (1996).
- Lekkerkerker, H. N. W. & Stroobants, A. Phase behaviour of rod-like colloid + flexible polymer mixtures. *Il Nuovo Cimento* **16 D** 949–962 (1994).
- Bohuijs, P., Stroobants, A., Frenkel, D. & Lekkerkerker, H. N. W. Numerical study of the phase behavior of rodlike colloids with attractive interactions. *J. Chem. Phys.* **107**, 1551–1564 (1997).
- Herzfeld, J. Entropically driven order in crowded solutions: From liquid crystals to cell biology. *Acc. Chem. Res.* **29**, 31–37 (1996).

26. Devanand, K. & Selser, J. C. Polyethylene oxide does not necessarily aggregate in water. *Nature* **343**, 739–741 (1990).
27. Model, P. & Russel, M. in *The Bacteriophages* (ed. Calender, R.) Ch. 6, 375–456 (Plenum, New York, 1988).
28. Chiruvolu, S., Naranjo, E. & Zasadzinski, J. A. *Microstructure of Complex Fluids by Electron Microscopy* (Am. Chem. Soc., Washington DC, 1994).

Acknowledgements. We acknowledge grant support from the NSF (D.M.R.).

Correspondence and requests for materials should be addressed to S.F. (e-mail: seth@smectic.else. brandeis.edu). Further information is available on the World Wide Web at <http://www.else.brandeis.edu>.

Low argon solubility in silicate melts at high pressure

Eva Chamorro-Pérez*, Philippe Gillet*, Albert Jambon†, James Badro* & Paul McMillan*‡

* *Institut Universitaire de France, Laboratoire de Science de la Terre, CNRS UMR 5570, École Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon cedex 07, France*

† *Laboratoire MAGIE, CNRS URA 1762, Université de Paris VI, 4 place Jussieu, 75252 Paris cedex 05, France*

The solubility of rare gases in silicate melts and minerals at high pressure is of importance for understanding the early history of the Earth and its present day degassing. Helium, neon, argon, krypton and xenon were originally incorporated into the Earth during its accretion, and have also been produced by radioactive decay¹. These elements have been used as tracers for deciphering mantle structure and constraining the number and size of geochemical reservoirs^{1–3}. In particular, it has been proposed that the budget of ⁴⁰Ar, produced by the radioactive decay of ⁴⁰K, provides the strongest argument for chemical layering within the mantle^{1,4}. The geochemical models used to arrive at this conclusion are, however, currently under re-examination⁵, with a large source of uncertainty being the lack of data on argon partitioning during melting. It has previously been assumed, on the basis of low-pressure data, that noble gases are highly soluble in melts at all pressures. But here we present solubility data of argon in olivine melt at very high pressure that indicate that argon solubility is strongly dependent on pressure, especially in the range of 4–5 gigapascals.

Degassing throughout Earth history has taken place by melting accompanied by dissolution of noble gases in the melt, followed by ascent of melt and escape to the atmosphere. Several recent studies have focused on rare-gas solubility in melts at low pressure (below 2.5 GPa)^{6–8}. Newly developed high-pressure diamond anvil cell (DAC) techniques now permit melting experiments on mantle materials in the presence of argon under deep-mantle conditions⁹. Preliminary experiments on pure SiO₂ melt yielded the surprising result that Ar solubility decreased markedly above 5 GPa. We have now obtained data for a natural olivine melt composition that provide a more realistic model for mantle melts.

Polished 100 × 100 × 20 μm³ plates (Fig. 1a) of single crystalline San Carlos olivine (chemical composition 49wt% MgO, 41wt% SiO₂, 10wt% FeO) were loaded into a high-pressure DAC along with Ar as pressure medium⁹. Pressure was measured from the calibrated shift of the intense Raman lines of the crystal before heating¹⁰. This avoided a potential chemical reaction with the ruby chips usually used for pressure measurement. Samples were heated at pressures up to 11 GPa by using a focused CO₂ laser beam¹¹. Forsterite melts congruently to at least 12.7 GPa (ref. 12). Melting was detected optically by the formation of droplets (Fig. 1b). The laser power was

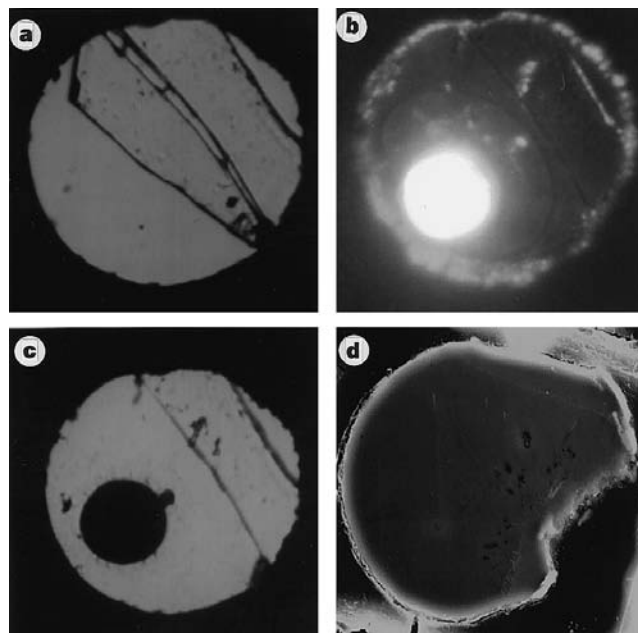


Figure 1 Series of optical micrographs taken through the diamonds of the DAC before and during the laser-heated experiment showing different stages of the high-pressure melting of olivine in the presence of Ar. **a**, Polished single-crystalline slab of San Carlos olivine compressed in solid Ar at 2.5 GPa before heating. The dark outline indicates the gasket hole in the diamond cell experiment. The largest dimension of the sample is 150 μm. The sample edges were fixed at two points of contact to the interior of the metal gasket. At ambient temperature and above 1.2 GPa, the Ar pressure-transmitting medium is solid. The line traversing the sample along its longest dimension is a fracture that occurred during compression. **b**, Photomicrograph of the sample *in situ* during laser heating at 2.5 GPa. The laser power was adjusted so that the temperature inside the molten spot was approximately 100–200 °C above the melting point. Only one piece of the fractured sample has melted to a spherical drop, and has moved away from the other half of the crystal (darker area above the melted spot). A rim of molten argon is visible around the hot bead of olivine melt: the Ar is solid elsewhere. **c**, Photomicrograph of the sample in the DAC after heating. When the laser power is cut, the molten sample quenches rapidly to glass (verified by polarized optical microscopy and micro-Raman spectroscopy, and by SEM analysis of the etched surface after removal from the DAC; see the text). The quenched bead of glassy sample droplet is then decompressed and prepared for chemical analysis. **d**, SEM image of a typical polished sample after etching (1 min in 1 M HCl). The sample surface shows only a smooth glassy texture. No fracture lines or bubbles are observed in the regions that were subjected to chemical analysis with the electron microprobe. The Ar content was measured by energy-dispersive and wavelength-dispersive analysis in the electron microprobe (Cameca SX50) with an accelerating voltage of 15 kV, a beam current of 100 nA and a counting time of 10 s on peak and background. The standard used was an SiO₂ glass containing 1wt% Ar. The Ar signal was measured simultaneously with two spectrometers. The electron beam was defocused to 5 μm in diameter to improve counting statistics on the low rare-gas content. The beam was rastered automatically across the sample surface in 5 μm steps. The reported Ar contents represent an average of approximately 20 points, depending on the sample analysed. A few empty 'bubbles' are visible on the right side of this sample: these are areas where the molten sample was 'pulled away' from the crystalline olivine during initial heating. These areas were avoided during the chemical analyses. All glassy samples analysed in this study formed monolithic blocks that remained intact during removal from the DAC, mounting in epoxy, and polishing. There was no evidence for microcracks or bubbles that might indicate decrepitation, for any of the glassy samples described here. Etching the sample surface did not reveal any additional texture that might indicate the presence of grain boundaries or other features associated with crystallization during the quench. Micro-Raman spectra of quenched samples occasionally showed the two most intense peaks of crystalline olivine superimposed on the broad bands of the glassy sample; however, these were not present for all samples, and there was no systematic difference between samples with high and low Ar content.

‡ Present address: Department of Chemistry and Biochemistry and Center for Solid State Science, Arizona State University, Tempe, Arizona 85287, USA.

adjusted to give temperatures just above the melting line (2,300–2,700 K)¹². During melting, the surrounding Ar melts by conduction, allowing the drop to move within the sample chamber (Fig. 1b). Samples were quenched by switching off the laser power, then removed from the DAC and polished for electron microprobe analysis (Fig. 1c). A conservative estimate with known thermal diffusivities of molten silicates gives the quench rate as at least 10^3 K s^{-1} . The actual quench rate is probably much higher, owing to radiative losses.

Chemical homogeneity was determined by elemental mapping at random points (>20) across the sample. In the regions analysed, no microcracks or argon-containing bubbles were detected by scanning electron microscopy (SEM) (Fig. 1d). Potential loss of Ar from the sample during quenching can be ruled out. Ar was found to be present in all the samples, and systematic chemical maps and profiles revealed no statistically significant variation of the Ar content from the core toward the edge of the sample, regardless of the pressure of the experiment. If diffusion of Ar had occurred one should have observed a systematic decrease in Ar concentration from the centre of the sample towards the edge, which did not occur. A further argument for no Ar loss is that thermal diffusivity was at least two orders of magnitude higher than chemical diffusivity (10^{-2} to $10^{-3} \text{ cm}^2 \text{ s}^{-1}$ compared with $10^{-5} \text{ cm}^2 \text{ s}^{-1}$). With our most conservative estimate of quenching time (10^{-3} s) one finds that the typical diffusion length of Ar is much less than $1 \mu\text{m}$, much smaller than the size of the sample.

The amount of incorporated argon increases smoothly with pressure up to 4–5 GPa, to reach a maximum of 0.2 wt%, as expected from an extrapolation of previous work on other melts at lower pressure^{6–8}. Beyond 5 GPa, the solubility decreases markedly, by an order of magnitude (Fig. 2). This behaviour is completely unexpected from previous studies on depolymerized melt compositions, but is consistent with our preliminary observation for SiO_2 (ref. 9). The surface textures of all samples analysed were identical, independently of run pressure. None of the samples revealed the presence of crystalline material by optical and SEM examination, even after etching (Fig. 1d). Weak features due to traces of forsterite were occasionally detected in micro-Raman spectra of quenched samples superimposed on broad glassy bands, but these were observed at all pressures, independently of Ar content. We conclude that our observations reflect a real change in Ar solubility in the melt at high pressure.

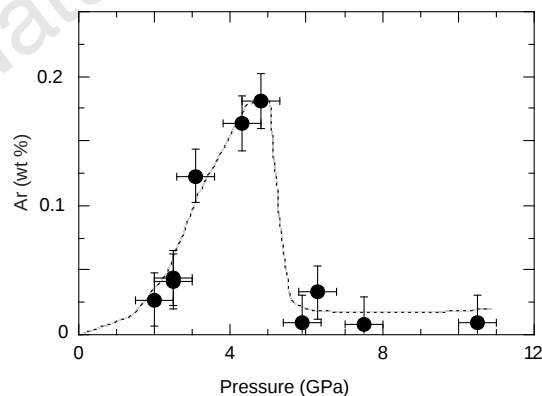


Figure 2 Argon solubility data for San Carlos olivine melt up to 10.5 GPa. Up to 4–5 GPa the solubility increases smoothly. Above this pressure a large decrease in solubility is observed. The dashed line does not represent a fit to the data, but is merely intended as a guide to the eye. It is obvious that a linear fit to these data would not yield a Henrian solubility behaviour passing through the origin, expected from previous work on high-silica melts at low pressure. This could indicate some deviation of the Ar solubility from ideality in the low-pressure range ($<3 \text{ GPa}$), or it could be within the error of the data. Further discussion of this point must await Ar solubility determinations at low pressure.

Previous rare-gas solubility measurements at low pressures have been used to construct thermodynamic models for the dissolution process^{8,13}. The models are based on the concept of ‘ionic porosity’, which describes the ‘free’ volume within the melt structure available to the chemically inert dissolving species, that is, the melt molar volume minus that occupied by the ions. The size distribution of available sites or ‘holes’ is determined from a series of solubility measurements with rare gases of different radii, and is described by the first moment of the distribution in a log–linear plot¹³. For a given melt, the hole size distribution is calculated from molar

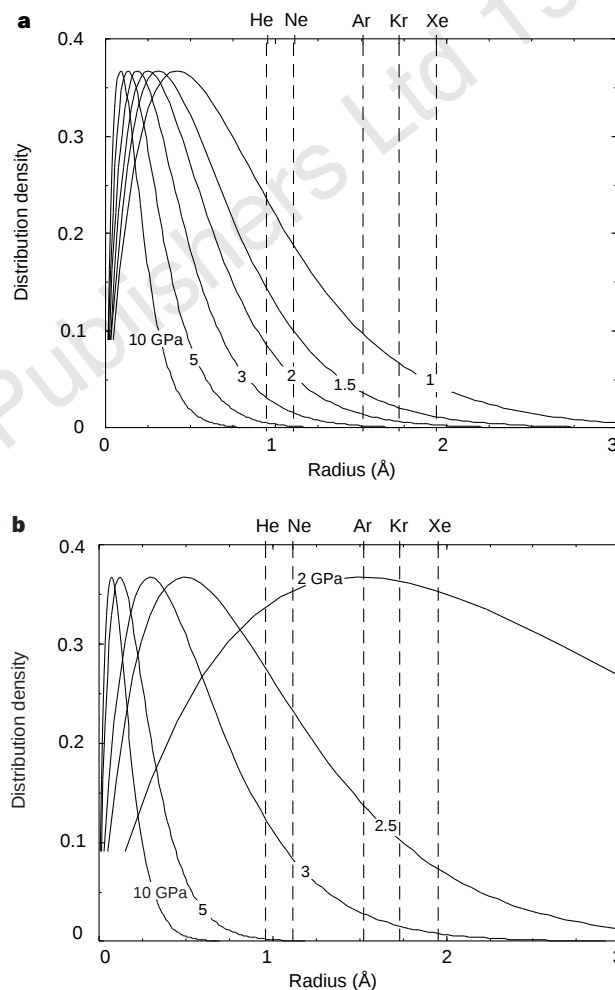


Figure 3 Hole size distribution in (a) olivine and (b) SiO_2 melts calculated with an ionic porosity model⁸. The distribution density is calculated with a log–linear model for site size probability¹³. The characteristic length coefficient is obtained by the slope of the measured solubility line against the atom radius for different rare gases. It is assumed that the decrease in ionic porosity and change in the hole size distribution is due only to a volume decrease of the melt with pressure, that is, that no major structural changes (such as coordination changes) occur in the melt over this pressure range. The change in volume for olivine melt is estimated with the following EOS parameters: $K_{\text{TO}} = 27 \text{ GPa}$; $K' = 5$; $dK/dT = -5 \times 10^{-3} \text{ GPa/K}$; $\alpha_0 = 7 \times 10^{-5} \text{ K}^{-1}$; ionic volume ($\text{Mg} + \text{Fe} + \text{Si} + \text{O}$) = $27.44 \text{ cm}^3/\text{mol}$, molar volume at $P_0/T_0 = 49.8 \text{ cm}^3/\text{mol}$ (ref. 15). The pressure dependence of the melting temperature of olivine has been included¹². For SiO_2 we used the EOS published by Poole *et al.*¹⁶. The radii of the rare gases He, Ne, Ar and Xe are indicated on the axis. For Ar, it is observed that the tail of the hole size distribution in olivine decreases abruptly to near zero at a calculated pressure of 3–5 GPa, in good agreement with the experimental solubility measurements. Below this pressure the available sites are not completely filled, and the solubility model of Carroll and Stolper⁸ applies. The analogous plot for SiO_2 is shown in (b). A similar interpretation holds for these data.

volume data and measured rare-gas solubilities. Over the pressure range considered by Carroll and Stolper⁸, it was assumed that the available sites were never completely filled; this model predicts a Henrian solubility behaviour. At higher pressure it would be expected that all the available sites would become filled, and the solubility would reach a maximum value. This prediction stands in marked contrast to the abrupt drop in Ar solubility above 5 GPa observed here. However, if changes in the ionic porosity or hole size distribution with pressure are taken into account, a rapid decrease in rare gas solubility above a certain threshold pressure would be expected. Even in the absence of major structural changes within the melt¹⁴, densification will cause a decrease in ionic porosity and change the distribution of the hole sizes (Fig. 3).

We have used an estimated equation of state (EOS) for molten olivine¹⁵ to calculate ionic porosities over the pressure range of our measurements. At low pressure, the available sites for Ar are not completely filled, and the effect of pressure is to increase the solubility as observed (Fig. 2). At higher pressure, the ionic porosity is lowered to such an extent that the tail of the distribution is now markedly decreased in the size range of the argon atoms (Fig. 3a). The result of these competing effects is an initial increase in

solubility with pressure, followed by a sudden decrease, as the range of available sites distribution crosses the Ar radius.

A similar argument can be used to explain the results previously observed for SiO₂ melt⁹ (Fig. 3b). Here we use a melt EOS obtained from molecular dynamics simulation¹⁶. The evolution of the pore size distribution with pressure is similar to that for olivine melt. This rationalizes why the Ar solubility decrease is observed at a similar pressure for both melts, although they are chemically and thermodynamically different. Comparison of the silica and olivine data raises the question of why the absolute Ar solubilities are so different for the two melts at low pressure. The maximum solubility for SiO₂ was 5wt% at 5 GPa, whereas that for molten olivine was 0.2wt% at the same pressure.

Silica is a fully polymerized electrically neutral framework that simply traps the neutral atoms within the melt cages. Olivine forms an ionic liquid, bonded by attractive and repulsive interactions between metal cations (for example Mg²⁺ or Fe²⁺) and SiO₄⁴⁻ anions. The average ionic separation should be close to that defined by the minimum-energy configuration. Placing Ar atoms in this liquid causes the ion-ion distance to increase, decreasing the net ionic attraction. Polarization of the rare-gas atoms will cause a further decrease in ion-ion attraction. The result is that the internal energy of the liquid is increased (in other words it becomes less negative). This 'ionic dilution' effect, which destabilizes the liquid thermodynamically, causes less argon to be dissolved in olivine melt than in silica.

This argument predicts that fully polymerized melts should have similar rare-gas solubilities to silica, once account is taken of available site filling by charge-balancing cations. The rare-gas solubility should decrease rapidly as the polymerization decreases. Preliminary data for enstatite and diopside melts support this argument. Ar solubilities in these melts are nearly identical to that in forsterite for pressures below 4 GPa (Fig. 4a). We predict that natural mantle melts have a similar behaviour, with an abrupt decrease in Ar solubility in the 3–5-GPa pressure range. The model can be used to predict the behaviour of other noble gases in mantle melts. It is expected that the solubility decrease would occur below 2 GPa for Xe but not until 8–10 GPa for He, owing to the different sizes of the noble gas atoms (Fig. 3). In our preliminary experiments⁹ we obtained one data point for anorthite melt at 10 GPa that did not agree with this model. But those experiments were plagued with problems associated with the incongruent melting of that composition at high pressure, and that solubility determination is unreliable. There are no such problems with silica and forsterite melts because their melting is congruent throughout the pressure range 0–10 GPa.

The sudden decrease in Ar solubility in the melt at 4–5 GPa has important consequences for its geochemical behaviour. It is assumed that noble gases are always strongly incompatible during partial melting, that is, that they are preferentially fractionated into the melt. This assumption is supported by low-pressure (<0.2 GPa) experimental data indicating that the olivine–melt partition coefficient ($K_d = [Ar]_{\text{crystal}}/[Ar]_{\text{melt}}$) is $\sim 10^{-2}$ (ref. 17). In previous geochemical modelling, it has been implicitly assumed that this value is independent of pressure. However, our data indicate an abrupt decrease in Ar solubility in the melt, by at least one order of magnitude, in the 4–5 GPa range. Unless there is a corresponding change in the crystal solubility at this pressure, which is unlikely, the behaviour of Ar will change from strongly incompatible toward moderately incompatible, or even compatible, if the melt solubility becomes lower than that of the crystal (Fig. 4b). Similar effects are expected at lower pressures for heavy noble gases such as Kr or Xe, and at higher pressure for He.

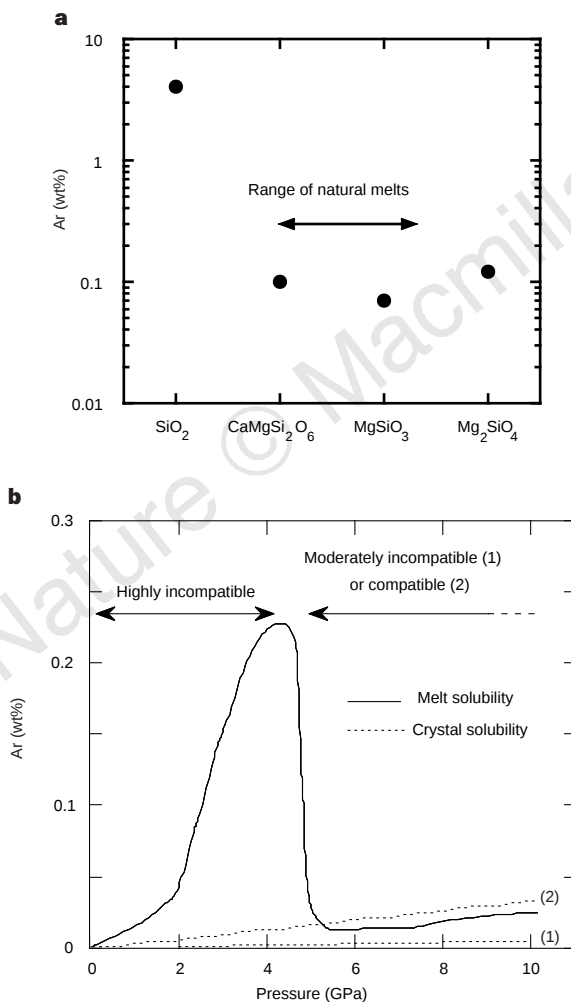


Figure 4 Solubility of argon in silicate melts as a function of composition and pressure. **a**, Argon solubility data at $P \approx 3$ GPa for SiO₂, diopside, enstatite and olivine melts. **b**, Comparison of Ar solubility in olivine melt measured as a function of pressure in this study with the expected solubility in crystalline olivine, obtained by extrapolating the low-pressure crystal–melt fractionation factor ($K_d \approx 10^{-2}$) to high pressure. Two possible Henry's-law crystal solubilities are shown, consistent with the low-pressure value of crystal–melt fractionation, within estimated limits of experimental error.

Received 27 June 1997; accepted 9 March 1998.

1. Jambon, A. in *Volatiles in Magmas* (eds Carroll, M. R. and Holloway, J. R.) 479–517 (Mineralogical Soc. Am., Washington DC, 1994).

2. Zhang, Y. & Zindler, A. Noble gas constraints on the evolution of the Earth's atmosphere. *J. Geophys. Res.* **94**, 13719–13737 (1989).
3. Azbel, I. & Tolstikhin, I. N. Geodynamics, magmatism, and degassing of the Earth. *Geochim. Cosmochim. Acta* **54**, 139–154 (1990).
4. Allègre, C., Hofmann, A. & O'Nions, K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23**, 3555–3557 (1996).
5. Albarède, F. Time-dependent models of U–Th–He and K–Ar evolution and the layering of mantle convection. *Chem. Geol.* (in the press).
6. White, B. S., Brearley, M. & Montana, A. Solubility of argon in silicate liquids at high pressures. *Am. Mineral.* **74**, 513–522 (1989).
7. Montana, A., Guo, Q., Boettcher, S., White, B. S. & Brearley, M. Xe and Ar in high pressure silicate liquids. *Am. Mineral.* **78**, 1135–1142 (1993).
8. Carroll, M. L. & Stolper, E. M. Noble gas solubilities in silicate melts and glasses: new experimental results for argon and the relationships between solubility and ionic porosity. *Geochim. Cosmochim. Acta* **75**, 5039–5051 (1993).
9. Chamorro-Perez, E., Gillet, P. & Jambon, A. Argon solubility in silicate melts at very high pressures. Experimental set-up and preliminary results for anorthite and silica melts. *Earth Planet. Sci. Lett.* **145**, 97–107 (1996).
10. Wang, S. Y., Sharma, S. K. & Cooney, T. F. Micro-Raman and infrared spectral study of forsterite under high pressure. *Am. Mineral.* **78**, 469–476 (1993).
11. Gillet, P., Fiquet, G., Daniel, I. & Reynard, B. Raman spectroscopy at mantle pressure and temperature conditions. Experimental set-up and the example of CaTiO₃ perovskite. *Geophys. Res. Lett.* **20**, 1931–1934 (1993).
12. Ohtani, E. & Kuazawa, M. Melting of forsterite Mg₂SiO₄ up to 15 GPa. *Phys. Earth Planet. Inter.* **27**, 32–38 (1981).
13. Shackelford, J. & Masaryk, J. The interstitial structure of vitreous silica. *J. Non-Cryst. Solids* **30**, 127–139 (1978).
14. Shibata, T., Takahashi, E. & Matsuda, J. Noble gas solubility in binary CaO–SiO₂ system. *Geophys. Res. Lett.* **23**, 3139–3142 (1996).
15. Agee, C. & Walker, D. Static compression and olivine flotation in ultrabasic silicate liquid. *J. Geophys. Res.* **93**, 3437–3449 (1988).
16. Poole, P. H., McMillan, P. F. & Wolf, G. H. in *Structure, Dynamics and Properties of Silicate Melts* (eds Stebbins, J. F., McMillan, P. F. & Dingwell, D. B.) 563–616 (Mineralogical Soc. Am., Washington DC, 1995).
17. Brooker, A., Wartho, J.-A., Carroll, M. & Kelley, P. Preliminary UVLAMP determinations of argon partition coefficients for olivine and clinopyroxene grown from silicate melts. *Chem. Geol.* (submitted).

Acknowledgements. We thank P. Blanc for his assistance with the SEM, D. Badia for help with microprobe analysis and K. Roselieb for giving the Ar–SiO₂ standard. We are grateful to R. A. Brooker for making his manuscript available before publication. This work was financially supported by the INSU program 'DBT: Terre Profonde'.

Correspondence and requests for materials should be addressed to P.G. (e-mail: pgillet@geologie.ens-lyon.fr).

Rare-gas solids in the Earth's deep interior

Andrew P. Jephcoat

Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK

Chemical inertness and surface volatility, combined with low abundance, have made the rare (noble) gases a unique trace-elemental and isotopic system for constraining the formation and evolution of the solid Earth and its atmosphere^{1–3}. Here I examine the implications of recent high-pressure measurements of the melting temperatures of heavy rare-gas solids—argon, krypton and xenon—with new diamond-anvil cell methods, together with their pressure–volume relationship, for the total rare-gas inventory of the Earth since its formation. The solid–liquid (melting) transition in these rare-gas solids rises significantly with pressure in the 50 GPa range^{4,5}, such that melting temperatures will exceed the geotherm at pressures of the Earth's transition zone and lower mantle (depths greater than 410–670 km). The densities of condensed rare-gas solids obtained from recent pressure–volume measurements at high compressions also exceed Earth's mantle and core densities. These pressure-induced changes in the physical properties of rare-gas solids, combined with their expected low solubilities and diffusional growth mechanisms, suggest that dense solid or fluid inclusions of rare gases—initially at nanometre scales—would have formed in the Earth's interior and may have resulted in incomplete planetary degassing. Separation of dense solid inclusions into deeper regions during early planet formation could provide a straightforward explanation for the

unexpectedly low absolute abundance of xenon observed in the atmospheres of both Earth and Mars.

All the rare gases crystallize under sufficient pressure to form close-packed solids that have been extensively studied with the diamond-anvil cell (DAC). Measured physical properties include the (300 K) equation of state (EOS; pressure–volume relationship), structural phase transitions, insulator–metal transitions and melting. Pressure–volume data to near 130 GPa at 300 K for solid Kr, reported here, complete the measurement of the EOS of the heavy rare gases (Ar, Kr, Xe) in this pressure range (Fig. 1). In addition, the solid–liquid phase transition boundary has been determined to pressures in the range 20–50 GPa (refs 4, 5) with the laser-heated DAC. Ar melting temperatures (T_m) coincide closely⁵ with several first-principles calculations to 50 GPa. If these melting temperatures are compared with a recent mantle geotherm based on experimental melting and phase relations in transition zone minerals⁶ (Fig. 2), they rise steeply above the geotherm at pressures (depths) around 8 GPa (270 km), 14 GPa (420 km) and 25 GPa (720 km) for Xe, Kr and Ar, respectively. Above these pressures, clusters of rare gas atoms not bound structurally within a host phase in the Earth would be thermodynamically stable in the solid phase. Confidence in the crossing of these melting curves with the geotherm is provided by the experimental observation of the coincidence of the melting curves of solid Ar and Kr with the melting curve of pure iron (Fig. 2) and solid Xe with platinum⁵. The melting curves to core pressures (130 GPa) are less well constrained, but the data indicate that all the rare-gas solids (RGS) should melt at higher temperatures than iron, the dominant core component. A softening of RGS melting curves (a less rapid increase with pressure) may mean that dense rare-gas fluids would be stable near the base of the lower mantle and in the core, depending on

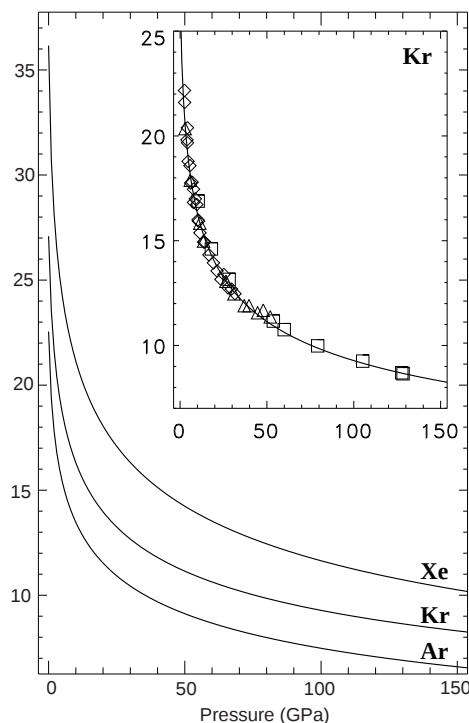


Figure 1 Molar volumes of solid Ar, Kr and Xe as a function of pressure at 300 K from DAC measurements. The solid lines are fits of an EOS using the condensed phase zero-pressure volume at low temperature (Table 1). The inset shows recent P – V data for solid krypton collected up to 128 GPa (squares) at Beamline X7A of the National Synchrotron Light Source, with angle-dispersive X-ray diffraction techniques plotted with earlier, independent measurements (triangles and diamonds).

(largely uncertain) core temperatures. Softening of the Xe melting curve below the insulator–metal transition above 100 GPa has been predicted⁷ such that its melting curve will cross that of solid Kr. Lower melting temperatures than those expected from semiempirical extrapolation of early measurements to 0.7 GPa for solid Xe are supported by the recent data⁵ to 12 GPa. Xe melting temperatures will always remain above the melting curve of iron for a positive Clapeyron slope of the transition.

Density can be calculated directly from the measured molar volumes with the normal (isotopically averaged) atomic weights of each rare gas ($\rho(P) = [\text{atomic weight}]/V_m(P)$). Figure 3 shows the 300 K density profiles obtained from the EOS of Fig. 1 and the expected density calculated at 2,500 K (representative of mantle temperatures) with a Mie–Grüneisen thermal model (Table 1) compared with the Earth's internal density⁸. The results show that the density of solid Ar (at 2,500 K) is approximately equal to the density in the lower mantle; solid Kr has approximately equal

density to core composition, and solid Xe is significantly denser than both the core and pure iron. (Higher temperatures in the core will produce only a small density reduction from that calculated at 2,500 K at these pressures.)

To determine whether dense clusters of rare gases (either solid or fluid) will form during cooling of planetary material containing rare gases, it is necessary to consider the solubility mechanisms at high pressures and assess the probability that discrete phases will form at the low concentrations observed in the Earth. Until recently, studies of the solubility of rare gases in melts⁹ have shown an increase with pressure and silica content up to ~ 1 GPa, but DAC work¹⁰ on the solubility of Ar in molten SiO_2 has suggested a rapid decrease in solubility above 5 GPa linked with structural changes in the melt. Shock-wave studies on the emplacement of volatiles into basalt¹¹ also indicate that emplacement efficiency drops off above ~ 40 GPa. Both types of study therefore suggest that the solubility of the rare gases in melts and solids falls as the host phases are compressed and

Table 1 Equation-of-state parameters for the heavy RGS

Parameter*	Argon	Krypton	Xenon
V_0 (cm ³ mol ⁻¹)	22.557	27.094	34.740
θ_0 (K)	93.3	71.9	57.0
γ_1	2.20	2.17	2.36
K_0 (GPa)	3.03(5)	3.32(5)	4.3(2)
K'_0	7.24(5)	7.23(5)	6.5(1)

* V_0 , zero-pressure volume obtained from low-temperature ($T \approx 4$ K) data; K_0 , bulk modulus, $-V(\partial P/\partial V)_T$; K'_0 , first pressure derivative of K_0 ; for definition of other parameters, see Fig. 3 legend.

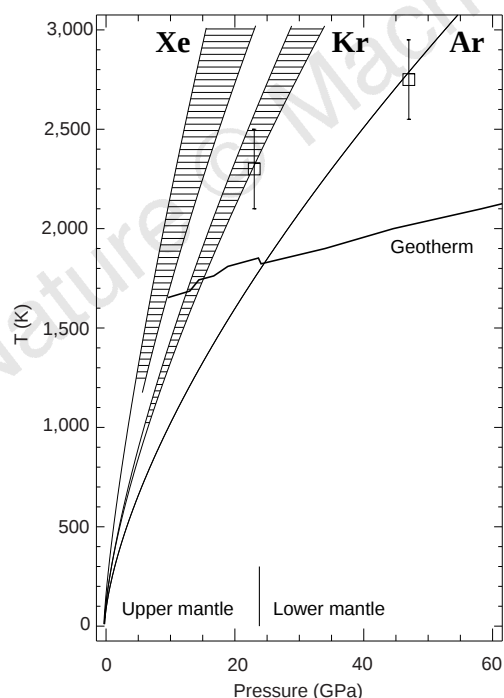


Figure 2 Expected melting curves of the heavy RGS based on data from high-pressure experiments and available theoretical models^{4,5}. Bounds on the melting curves of solid Kr and Xe are shown by the shaded regions. The two symbols with error bars indicate the P – T region where overlap with the melting temperature of solid iron has been observed: solid Kr is stable in an iron melt above ~ 23 GPa and 2,300 K, and solid Ar is stable above ~ 47 GPa and 2,700 K. The geotherm is obtained from thermodynamic data on mantle phases⁶ and increases adiabatically with depth (bold line). The temperature of the lower mantle falls well below the expected melting temperatures of the RGS.

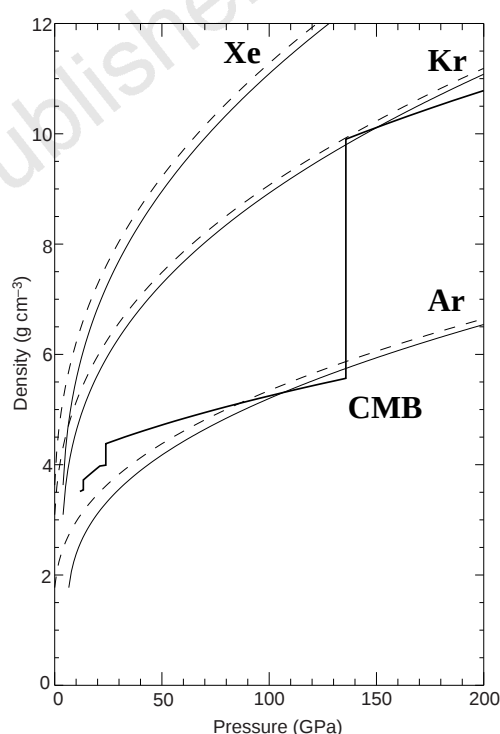


Figure 3 Pressure–density data on the heavy RGS extended to 200 GPa evaluated with normal atomic weights for each element. Dashed lines are densities calculated from the experimental data at room temperature of Fig. 1. The lowering in the density expected at 2,500 K, calculated with the Mie–Grüneisen thermal model and the Debye approximation, is shown by solid lines for each rare-gas solid: the room-temperature compression data contain both zero-point (P_{ZP}) and thermal-pressure (P_{TH}) contributions to the total pressure ($P(V, T) = P_{SL}(V) + P_{ZP}(V) + P_{TH}(V, T)$, where $P_{SL}(V)$ is the static-lattice term at $T = 0$); P_{ZP} and P_{TH} are functions of the Grüneisen parameter and Debye temperature. P_{TH} as a function of volume can be estimated using the data listed in Table 1, if the volume dependence of the Grüneisen parameter, γ , is assumed to be linearly proportional to the relative volume: $\gamma = \gamma_1(V/V_0) + \gamma_0$, with γ_0 fixed at 0.5. The P – V relations at higher temperatures can then be calculated by adding this thermal-pressure term (in this case, at 2,500 K) to the 0 K isotherm for each RGS. It is significant that the density difference between 300 K and 2,500 K isotherms decreases with increasing pressure. The calculations do not include any temperature dependence of the Debye temperature, θ_D , (in addition to its volume dependence) nor anharmonic effects at these high temperatures, but such corrections are expected to be small. Superposed on the RGS densities is the density profile of the Earth obtained from a seismological model⁸ (see text). The sharp jump in density marks the core–mantle boundary (CMB) pressure.

the available free volume decreases, in accord with an ionic porosity model⁹. Further, ion implantation experiments in both metals and ceramics¹² show that atomic-scale precipitates and microscopic bubbles of all rare gases form readily in conditions of supersaturation (low solubility). These inclusions have high internal pressures—solid Kr bubbles were formed epitaxially¹³ in Ni, for example, at pressures as high as 6 GPa and melted at temperatures consistent with the recent laser-heated DAC data⁵. (High bubble pressures are maintained by a balance with surface energies of the host at small radii.)

An estimate of the coagulation rate of rare-gas atoms can be calculated from the clustering probability of particles derived for colloidal solutions based on a brownian-motion model¹⁴, with rare gas diffusion coefficients taken from low-pressure measurements¹⁵. For a (worst-case) estimate of the concentration of Xe in diamond¹⁶ of $10^{-11} \text{ cm}^3 \text{ STP g}^{-1} \equiv 10^8 \text{ Xe atoms cm}^{-3}$ (representative of mantle concentration), accumulation of the order 200 Xe atoms, sufficient to form a nanometer-size macroscopic solid¹², can occur on a timescale of less than a year. Significantly larger particles may therefore be expected to grow over the timescale of Earth evolution: growth of rare-gas bubbles in implanted solids is known to be promoted by vacancy trapping and thermal vacancy production at grain boundaries at temperatures $\geq 0.5T_m$ —conditions that would have been dominant since planet formation through to the present—and enhanced by migration and coalescence growth mechanisms¹². Formation of rare-gas inclusions in the Earth may then be a natural consequence of their low solubility and inertness. Furthermore, recent calculations show no tendency for alloy formation between Xe and iron at Earth's core conditions¹⁷, and support the hypothesis made here that phase separation should dominate over chemical bonding or solution at 100-GPa pressures.

The occlusion of rare gases in planetary interiors can take place through initial adsorption on solid surfaces of accreting material². Other models of planetary accretion suggest a dense proto-atmosphere may have further increased the solubility of rare gases into the surface of a magma ocean¹⁸. Convection of this material into deeper regions would have enhanced the probability of cluster formation as the solubility drops. RGS will then form successively in a cooling planetary body as the internal temperature crosses the respective melting temperatures (Fig. 2). Xe would have precipitated first to form inclusions with highest density, followed by Kr and Ar. In a solid mantle, these inclusions would remain locked as passive tracers. Processes of planetary formation and differentiation are complex and include models of crystal settling in magmas, with widely varying degrees of fractional crystallization according to size of the magma body¹⁹, as well as the need for percolation of dense iron alloy melts through surrounding silicate during Earth's core formation²⁰. Because growth of rare gas clusters is geologically fast (even at the low absolute concentrations of Xe), the question arises whether the comparable density of (fluid or solid) inclusions of the rare gases to Earth materials could have driven separation into deeper regions in either of these planetary differentiation stages, depending on relative melting temperature of the host material. There are regimes in unsteady convection where the effective friction velocity at boundary layers may drop sufficiently to allow settling of particles in the flow²¹ which may take place in silicate melts with low viscosity. Irrespective of the density, mechanical transport of RGS particulates in iron into the planetary core may have been sufficient to establish a separate core reservoir. Calculations of re-entrainment from a bed of particles into a convecting system²² suggest that the crystal sizes above which re-entrainment is not readily achieved are surprisingly small—especially for the Earth's core, where particle sizes must remain below $\sim 10^{-8} \text{ m}$ to be re-entrained from the core boundaries. As estimated above, the growth of RGS inclusions could well exceed this size over the age of the Earth.

If growth of dense inclusions modified the distribution of rare gases during Earth formation, it has a number of implications for the evolution of the Earth's atmosphere (and planetary atmospheres in general). In particular, it can provide a straightforward argument for the relative depletion of Xe required by models of the mass fractionation of Xe isotopes in the terrestrial planets: models of hydrodynamic escape^{2,23} or gravitational capture of solar gas into planetesimals^{24,25} can reproduce the Xe isotope fractionation observed in present-day atmospheres, but only for much higher Xe total abundance relative to other rare gases. Mechanisms proposed for preferentially lowering the abundance of Xe on both Earth and Mars include early outgassing followed by hydrodynamic escape^{26,27} or storage deep in the planet by chemical partitioning into metallic (core) phases^{2,23}. The partitioning of Xe into core phases was originally thought to be driven by the metallic transition in Xe expected near 130 GPa as seemed likely on Earth, but the low central pressure on Mars ($\sim 40 \text{ GPa}$) precluded this argument², as do first-principles simulations of bonding in the Xe-Fe system¹⁷. On the other hand, martian pressures are sufficient to raise the melting temperatures of Kr and Xe into the stability field of the solid phases at planetary temperatures, opening the possibility for mechanical separation of the rare gases and reviving the deep-storage hypothesis. Catastrophic degassing of the heavy rare gases need not have occurred during a rapid accretion phase of the Earth, because the rise in central pressures would effectively occlude RGS into both solid and melt fractions of host material such that the terrestrial planets may well be incompletely degassed, a hypothesis that conflicts with the assumption of many models of planetary atmospheric rare-gas evolution¹⁶. To extend the hypothesis, the lower internal pressures acting during the formation of Mars may have been less effective in trapping solid Ar than on Earth, and, if martian internal temperatures are now as low as 2,000 K, as some models suggest, what Ar does exist in its interior, as well as Kr and Xe, will have precipitated as solids isolated from the surface after cessation of tectonics.

The natural radiogenic production of rare-gas isotopes provides a mechanism for the formation of rare-gas inclusions directly at depth, analogous to ion implantation into host materials at room pressure, and is of most interest in the Earth's lower mantle and core where the effect of pressure on RGS properties should have been most pronounced. ^{40}Ar has been produced in both the lower mantle³ and probably the core²⁸ throughout the age of the Earth from radioactive decay of ^{40}K . Applying the melting and diffusion arguments described above, ^{40}Ar formed in the core would accumulate in the solid phase with a density of roughly 50% of that of the outer core. Migration to, and coalescence at the top of, the outer core, where it would be insoluble in lower-mantle material, with approximately equal density, could hinder incorporation into the mantle and represent a reservoir of ^{40}Ar isolated from the Earth's surface and the atmosphere since core formation, a factor not accounted for in mantle evolution models³. The quantities of ^{40}Ar available are much greater than for Kr and Xe isotopes: if all the ^{40}Ar postulated to remain in the Earth³ had collected at the core-mantle boundary, it would form a shell $\sim 0.08 \text{ m}$ thick at core conditions. Isolated particles of ^{40}Ar would readily exceed the critical dimension for re-entrainment by convection deeper into the core.

Radiogenic xenon isotope systematics have been used to constrain the age of the Earth relative to chondritic meteorite ages. These estimates are based on abundance ratios of ^{129}I and ^{244}Pu , which have distinct timescales for decay to the various Xe isotopes¹⁶. The effect of pressure on chemical partitioning of Xe parent nuclides ^{129}I or ^{244}Pu remains to be explored in the 100-GPa range, but if I and Pu behaved differently under pressure, such that ^{129}I partitioned into the Earth's core more effectively

than Pu, solidification and isolation of the ^{129}Xe produced in the core could distort the Earth's surface Xe radiogenic isotope signatures. Xe particles settling at the inner core boundary may then also exceed the size for re-entrainment through growth mechanisms²². In general, pressure effects in planetary bodies during accretion may go some way towards explaining the underabundance of ^{129}Xe in the Earth's atmosphere¹⁶, explain why radiogenic heavy isotopes appear in meteorites but are less apparent on Earth and Mars, and modify the 'delayed-formation' hypothesis for the Earth^{16,29}. The arguments here have used pure, end-member properties to describe each rare gas separately. If diffusion processes mix the rare gases, as implantation experiments suggest at normal pressure, then He or Ne trapped in stable, stoichiometric binary alloys with the heavy rare gases at high pressures (of the type formed in hard-sphere mixtures³⁰) could also have been occluded in the deep Earth. Formation of a stoichiometric Xe–Ne alloy, for example, may explain the presence of solar-type abundance patterns for Ne in plume sources like the Loihi seamount, Hawaii³¹, and thought to originate from the lower mantle. □

Received 20 March; accepted 9 April 1998.

- Ozima, M. Ar isotopes and Earth–atmosphere evolution models. *Geochim. Cosmochim. Acta* **39**, 1127–1134 (1975).
- Pepin, R. O. On the origin and early evolution of terrestrial planet atmospheres and meteoritic volatiles. *Icarus* **92**, 2–79 (1991).
- Allègre, C. J., Hofmann, A. & O'Nions, R. K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23**, 3555–3557 (1996).
- Jephcoat, A. P. & Besedin, S. P. Temperature measurement and melting determination in the laser-heated diamond-anvil cell. *Phil. Trans. R. Soc. A* **354**, 1333–1360 (1996).
- Jephcoat, A. P. & Besedin, S. P. in *Properties of Earth and Planetary Materials at High Pressure and Temperature* (eds Manghni, M. H. & Yagi, T.) 287–296 (Geophys. Monogr. 101, Am. Geophys. Union, Washington DC, 1998).
- Akagi, M. & Ito, E. Heat capacity of MgSiO_3 perovskite. *Geophys. Res. Lett.* **20**, 105–108 (1993).
- Young, D. A. *Phase Diagrams of the Elements* (University of California Press, Berkeley, 1991).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
- Carroll, M. R. & Stolper, E. M. Noble gas solubilities in silicate melts and glasses: New experimental results for argon and the relationship between solubility and ionic porosity. *Geochim. Cosmochim. Acta* **57**, 5039–5051 (1993).
- Chamorro-Perez, E., Gillet, P. & Jambon, A. Argon solubility in silicate melts at very high pressures. Experimental set-up and preliminary results for trapped gases in shergottite EETA 79001. *Geochim. Cosmochim. Acta* **52**, 295–307 (1988).
- Wiens, R. C. & Pepin, R. O. Laboratory shock emplacement of noble gases, nitrogen, and carbon dioxide into basalt, and implications for trapped gases in shergottite EETA 79001. *Geochim. Cosmochim. Acta* **52**, 295–307 (1988).
- Templier, C. in *Fundamental Aspects of Inert Gases in Solids* (eds Donnelly, S. E. & Evans, J. H.) 117–132 (NATO ASI B279, Plenum, New York, 1991).
- Evans, J. H. & Mazey, D. J. Evidence for solid krypton bubbles in copper, nickel, and gold at 293 K. *J. Phys. F: Met. Phys.* **15**, L1–L6 (1985).
- Chandrasekhar, S. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **15**, 1–89 (1943).
- Lux, G. The behavior of noble gases in silicate liquids: Solution, diffusion, bubbles and surface effects, with applications to natural samples. *Geochim. Cosmochim. Acta* **51**, 1549–1560 (1987).
- Ozima, M. & Podosek, F. A. *Noble Gas Geochemistry* (Cambridge University Press, UK, 1983).
- Caldwell, W. A. *et al.* Structure, bonding and geochemistry of xenon at high pressures. *Science* **227**, 930–933 (1997).
- Harper, C. L. & Jacobsen, S. B. Noble gases and Earth's accretion. *Science* **273**, 1814–1818 (1996).
- Tonks, W. B. & Melosh, H. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 151–174 (Oxford University Press, Oxford, 1990).
- Stevenson, D. J. Models of the Earth's core. *Science* **214**, 611–619 (1981).
- Martin, D. & Nokes, R. Crystal settling in a vigorously convecting magma chamber. *Nature* **332**, 534–536 (1988).
- Solomatov, V. S., Olson, P. & Stevenson, D. J. Entrainment from a bed of particles by thermal convection. *Earth Planet. Sci. Lett.* **120**, 387–393 (1993).
- Pepin, R. O. Evolution of Earth's noble gases: consequences of assuming hydrodynamic loss driven by giant impact. *Icarus* **126**, 148–156 (1997).
- Ozima, M. & Zahnle, K. Mantle degassing and atmospheric evolution: Noble gas view. *Geochem. J.* **27**, 185–200 (1993).
- Ozima, M., Wieler, R., Marty, B. & Podosek, F. A. Comparative studies of solar, Q-gases and terrestrial noble gases, and implications on the evolution of the solar nebula. *Geochim. Cosmochim. Acta* **62**, 301–314 (1998).
- Tolstikhin, I. N. & O'Nions, R. K. The Earth's missing xenon: A combination of early degassing and of rare gas loss from the atmosphere. *Chem. Geol.* **115**, 1–6 (1994).
- Zhang, Y. & Zindler, A. Noble gas constraints on the evolution of the Earth's atmosphere. *J. Geophys. Res.* **94**, 13719–13737 (1989).
- Ito, E., Morooka, K. & Ujike, O. Dissolution of K in molten iron at high pressure and temperature. *Geophys. Res. Lett.* **20**, 1651–1654 (1993).
- Wetherill, G. W. Radiometric chronology of the early solar system. *Annu. Rev. Nucl. Sci.* **25**, 283–328 (1975).
- Eldridge, M. D., Madden, P. A. & Frenkel, D. Entropy-driven formation of a superlattice in a hard-sphere binary mixture. *Nature* **365**, 35–37 (1993).
- Podosek, F. A. Solar gases in the Earth? *Nature* **349**, 106–107 (1991).

Acknowledgements. I thank R. O. Pepin and M. Ozima for comments and discussion.

Optical and radiocarbon dating at Jinmium rock shelter in northern Australia

Richard Roberts, Michael Bird*, Jon Olley†, Rex Galbraith‡, Ewan Lawson§, Geoff Laslett||, Hiroyuki Yoshida, Rhys Jones¶, Richard Fullagar#, Geraldine Jacobsen§ & Quan Hua§

Department of Earth Sciences, La Trobe University, Melbourne, Victoria 3083, Australia

* Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 0200, Australia

† CSIRO Land & Water, Canberra, Australian Capital Territory 2601, Australia

‡ Department of Statistical Science, University College London, London WC1E 6BT, UK

§ Physics Division, Australian Nuclear Science and Technology Organisation, Menai, New South Wales 2234, Australia

|| CSIRO Mathematical & Information Sciences, Melbourne, Victoria 3168, Australia

¶ Department of Archaeology & Natural History, RSPAS, Australian National University, Canberra, Australian Capital Territory 0200, Australia

Division of Anthropology, Australian Museum, Sydney, New South Wales 2000, Australia

The Jinmium rock shelter is located in the Kimberley region of northern Australia. Claims for ancient rock art and an early human presence at this site¹ were based on thermoluminescence ages of 50–75 thousand years (kyr) for quartz sands associated with buried circular engravings (pecked cupules) and on thermoluminescence ages of 116–176 kyr for the underlying artefact-bearing deposits. Here we report substantially younger optical ages for quartz sand, and ages based on measurements of radioactive carbon in charcoal fragments, from the occupation deposit. Using conventional (multiple-grain) optical dating methods, we estimate that the base of the deposit is 22 kyr. However, dating of individual grains shows that some have been buried more recently. The single-grain optical ages indicate that the Jinmium deposit is younger than 10 kyr. This result is in agreement with the late-Holocene ages obtained for the upper two-thirds of the deposit from radiocarbon measurements. We suggest that some grains have older optical ages because they received insufficient exposure to sunlight before burial. The presence of such grains in a sample will cause age overestimates using multiple-grain methods, whether using thermoluminescence or optical dating.

The Jinmium rock shelter is formed under a small tilted block of sandstone. Excavation of the floor deposits began in 1993 and thermoluminescence ages¹ were obtained from quartz grains of 90–125 μm in diameter that constitute 3–5% (by mass) of the unconsolidated sandy sediments. The deposit has a maximum depth of 2 m, with abundant boulders and friable rubble below a depth of 55 cm (Fig. 1). Pecked cupules cover the northwest wall of the shelter and were traced below ground level to a maximum depth of 97 cm; an engraved sandstone fragment was recovered from a depth of 100 cm. The lowest artefacts were found at a depth of 160 cm (30 cm below the 116 kyr thermoluminescence sample) and ochres were collected from slightly higher levels¹. Charcoal pieces were collected by the original excavators in 1993, from deposits sieved through 2-mm and 4-mm meshes, and accelerator mass spectrometry (AMS) measurements of ^{14}C levels were made on these samples after acid–alkali–acid pretreatment (Table 1).

A fresh set of sediment samples for optical and ^{14}C dating was collected in 1995 (by R.R., R.J. and R.F.) from the cleaned faces of the original pit. We collected three ^{14}C samples from the north face of the excavation, and two from the south face (Fig. 1). Charcoal

fragments larger than 125 μm were hand-picked from the wet-sieved sediment samples, and charcoal pieces of $<125\text{ }\mu\text{m}$ were isolated by chemical removal of silicates. We then decontaminated both size fractions (methods adapted from ref. 2) to extract the elemental carbon. AMS determinations³ were made on the fractions consisting of particles of $>125\text{ }\mu\text{m}$ from four samples, the finer fraction of two samples, and (because of the paucity of material) the combined mass of these two fractions for the deepest sample

(Table 1). To check that young carbon was not introduced during sample preparation, we prepared a charcoal sample from Ngarra-bullgan Cave⁴ (north Queensland) in an identical way. This sample (OZC733) gave an AMS age of $35.5 \pm 0.6\text{ kyr BP}$, in agreement with the similar ^{14}C -based ages, in the range 29–36 kyr BP, obtained previously from 20 charcoal samples from the same stratum⁴.

Thermoluminescence and optical dating^{5–8} provide an estimate of the time since minerals such as quartz were last exposed to

Table 1 Accelerator mass spectrometry ^{14}C -based ages for Jinmium

Sample code	Field code	Depth* (cm)	Mesh size or fragment size†	Conventional age‡ (years BP)
Carbon samples collected in 1993				
SUA-3036§	C1/I/10	70	–	$1,790 \pm 150$
OZA306	C1/I/12	85–90	2–4 mm	370 ± 50
OZA307	C1/I/13	90–95	2–4 mm	$2,100 \pm 60$
Beta-81816	C1/I/15	120	2–4 mm	950 ± 60
OZA303	C1/II/21	60	2–4 mm	80 ± 50
Beta-81817	C1/II/23A	100	2–4 mm	$1,860 \pm 60$
OZA304	C1/III/14A	85–90	2–4 + > 4 mm	$2,800 \pm 60$
OZA305	C1/III/22B	115	2–4 mm	860 ± 40
OZA302	C1/III/24A	130	2–4 mm	$3,870 \pm 70$
Elemental carbon samples collected in 1995				
North face				
OZC605	COOR 1/3	68	>125 μm	$1,080 \pm 90$
OZC539		(C1/III)	<125 μm	$1,130 \pm 110$
OZC603	COOR 1/2	71	>125 μm	$1,020 \pm 100$
OZC538		(C1/II)	<125 μm	$1,760 \pm 90$
OZC602	COOR 2/2	110	>125 μm	$2,890 \pm 90$
		(C1/III)		
South face				
OZC604	COOR 8/1	96	>125 μm	$3,330 \pm 100$
OZC763	COOR 6/2	142	< 125 + > 125 μm	$1,100 \pm 60$
		(C1/III)		

* Depths for samples collected in 1993 are estimates based on locations within excavation units.
† Mesh size (in mm) for samples collected in 1993. Fragment size (in μm) for samples collected in 1995.
‡ Mean $\pm 1\sigma$ uncertainty. Calibration²⁰ decreases the 860–2,100 years BP ages by 50–140 years and increases the 2,800–3,870 years BP ages by 70–410 years. A $\delta^{13}\text{C}$ value of -25.9‰ was measured for Beta samples and a value of -25‰ was assumed for OZ and SUA samples.
§ Radiometric ^{14}C -based determination cited in ref. 1. Sample was hand-picked from section.

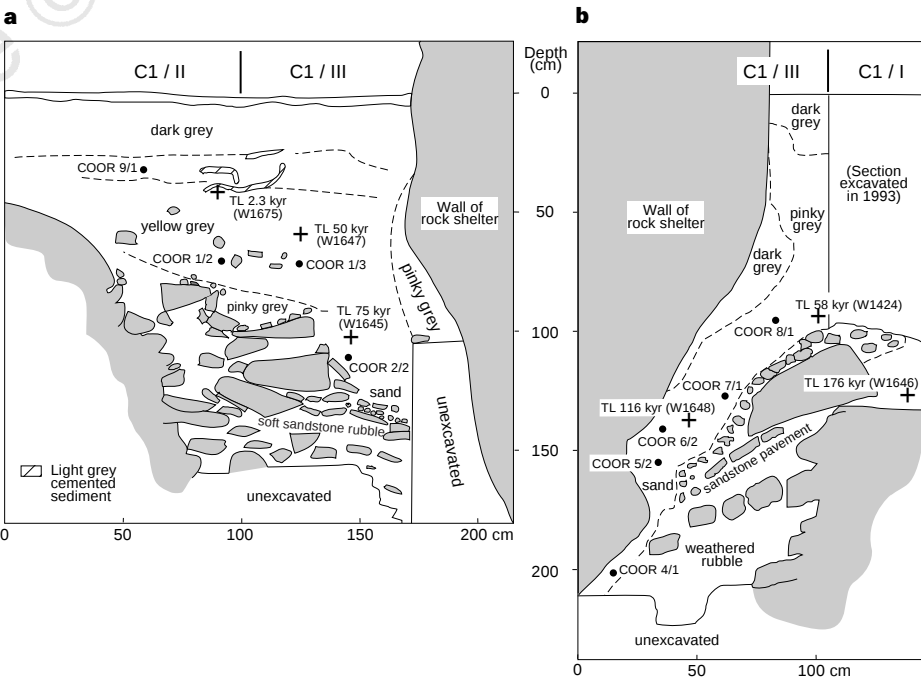


Figure 1 Stratigraphic sections of north and south faces of the Jinmium excavation. **a**, Stratigraphic section of Jinmium pits C1/II and C1/III (north faces), drawn during the 1995 re-excavation. Locations of optical and elemental carbon ^{14}C samples are shown by filled circles and the COOR field codes.

Thermoluminescence (TL) sample locations are marked by crosses, together with mean ages and Wollongong laboratory codes¹ (in parentheses). **b**, Stratigraphic section of Jinmium pit C1/III (south face), drawn during sample collection in 1995. Optical, ^{14}C and thermoluminescence samples are labelled as in **a**.

sunlight. Buried grains will accumulate the effects of the nuclear radiation flux to which they are exposed, and the burial dose (palaeodose) can be measured using thermoluminescence or optically stimulated luminescence (OSL). Weathering of buried bedrock and rubble will release grains that have not been exposed to sunlight since the time of rock formation, so the presence of such grains in a deposit and their analysis by multiple-grain methods will give falsely old thermoluminescence and optical ages. Grains exposed to insufficient sunlight before burial will likewise give rise to age overestimates; optical dating is better suited than thermoluminescence dating to such deposits because the OSL signal is more rapidly reset by sunlight⁹. Furthermore, optical dating of single grains of feldspar^{10,11} or quartz^{12,13} allows insufficiently bleached grains to be detected and thus excluded from the age calculation. We chose eight samples for optical dating using conventional multiple-aliquot additive-dose methods, three samples from the north face of the

1995 re-excavation and five from the south face (Fig. 1 and Table 2). We also dated six of these samples using regenerative-dose protocols devised for single aliquots^{12–15}; we first used aliquots composed of many (~800) grains and then examined more than 1,000 grains individually.

We consider that the elemental carbon samples of >125 μm (Table 1) yield the most reliable ^{14}C -based ages. Their stratigraphic locations are known with greatest precision; the coarser fraction is the least likely to suffer from significant postdepositional movement; and the decontamination pretreatment² removes any organic carbon components that may not be completely removed by conventional pretreatments. The ^{14}C -based ages for these samples exhibit good stratigraphic coherence and yield ages of about 1.0 kyr BP and 3.0 kyr BP for the deposits at ~70-cm and ~100-cm depth, respectively. Less consistency is shown among ^{14}C -based ages obtained from the <125 μm and mixed-size fractions, and from

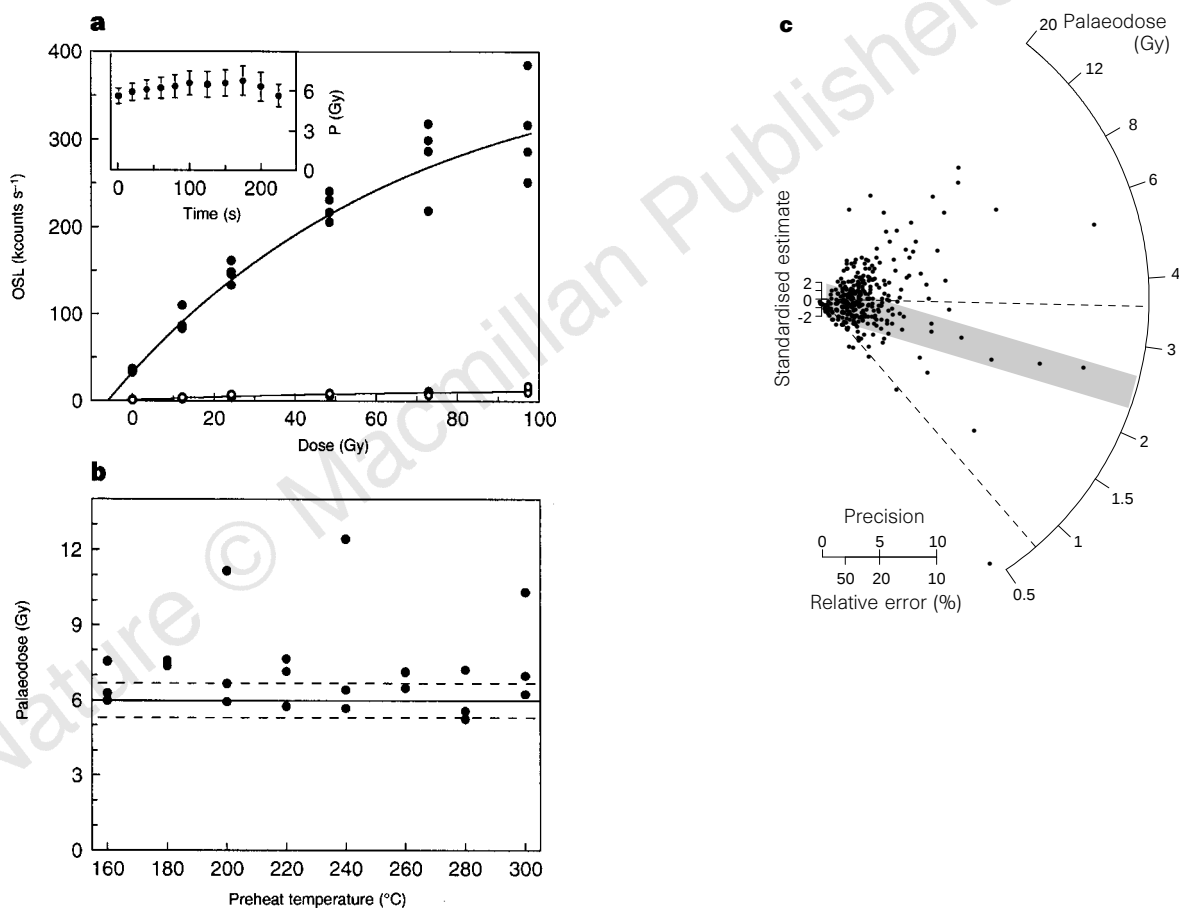


Figure 2 Optical dating of sample COOR 8/1. **a**, OSL additive-dose growth curve (upper line) fitted to 24 multiple-grain (~3-mg) aliquots (filled circles) of sample COOR 8/1. The thermal-transfer correction²¹ curve (lower line) is fitted to another set of 24 aliquots (open circles) which have been bleached for 250 s with light of 420–550 nm in wavelength before preheating. The intercept of the two curves gives the palaeodose (6.0 ± 0.7 Gy); the upper curve intercepts the baseline at a similar point, indicating that thermal transfer was insignificant for this sample. The inset shows the plot of palaeodose (P) versus illumination time. **b**, OSL palaeodoses obtained from sample COOR 8/1 at different preheat temperatures using a regenerative-dose protocol applied to 24 single aliquots, each composed of multiple (~1-mg) grains^{14,15}. Three aliquots were held for 10 s at each preheat temperature. The error in palaeodose for each aliquot is smaller than the size of the filled circle. The plateau from 160 °C to 300 °C implies negligible thermal transfer. The mean multiple-aliquot palaeodose determined in **a**, and its standard error, are marked by solid and dashed lines, respectively; the three aliquots with significantly larger palaeodoses presumably contain a greater proportion of

insufficiently bleached grains. **c**, Radial plot^{29,30} of palaeodoses determined for 322 individual grains from sample COOR 8/1 using a regenerative-dose single-aliquot OSL protocol^{14,15}. Single grains vary in OSL intensity, resulting in differing palaeodose precisions. The y-axis plots standardised estimates of log palaeodose (log palaeodose minus the weighted average for all grains, divided by the standard error of log palaeodose). The x-axis (precision) plots reciprocal standard errors of log palaeodose; these may be converted to approximate relative errors of palaeodose, which are also shown. Points on any straight line radiating from the origin have the same log palaeodose; a circular scale has been added so that palaeodoses (log scale) may be read off. Palaeodoses range from <0.5 to >20 Gy. Points that are statistically concordant will mostly lie within ± 2 units (on the y-axis) of a common radial line, as shown by the shaded component centred at 2.5 Gy. This palaeodose yields an optical age in agreement with the ^{14}C -based chronology; it is comfortably bracketed by the estimates obtained from the central-age and minimum-age models shown by the upper and lower dashed lines, respectively.

field-sieved charcoal; ^{14}C -based ages for the latter samples range from 0.1 to 3.9 kyr BP and are not in coherent stratigraphic order, although the deepest sample (located just above the 116 kyr thermoluminescence sample) yields the oldest age. Extraneous material may have been introduced into some of the 1993 samples during excavation or sieving. Mixing of the deposit through physical disturbance (for example, by soil fauna) may also account for the age of 1.1 kyr BP given by mixed-size elemental carbon at a depth of 142 cm.

Multiple-aliquot and single-aliquot (multiple-grain) optical dating methods yield similar ages for the same samples (Table 2 and Fig. 2a, b). Ages range from 2.2 kyr near the surface to 22 kyr at 2-m depth, and are in correct stratigraphic order, but both sets of ages are too old compared with the chronology for elemental carbon samples of $>125\text{ }\mu\text{m}$ in the top half of the deposit. We view these optical ages as maximum estimates, because each aliquot may include grains that were insufficiently bleached before burial. The presence of such grains in the deposit is shown directly by optical dating of individual sand grains (Table 2 and Fig. 2c). (Palaeodoses in the range of 0–10 Gy are calculated with greatest accuracy, whereas those above 20 Gy greatly underestimate the true palaeodose because of the nonlinear dose-response characteristics of quartz and the method of palaeodose determination used here.) Dose-response curves constructed for single grains with palaeodoses of $>20\text{ Gy}$ show that some grains from sample COOR 8/1 and

deeper samples are in dose saturation, a condition consistent with being derived from weathered rubble or bedrock. Each of the six samples also exhibits a spread in single-grain palaeodoses with values of $<20\text{ Gy}$, rather than a single common palaeodose. We interpret the spread as resulting from insufficient exposure to sunlight of many of the grains before burial. Under these conditions, grains with the smallest palaeodoses yield the more reliable age estimates, as they are apt to have had sufficient exposure to sunlight. We have accordingly calculated optical ages for the six samples using a modified minimum-age model¹⁶ (G2 values in Table 2).

These minimum ages are probably too young, however, because of the inclusion of grains that have intruded from overlying (younger) levels and grains whose beta-particle dose rate is greater than the sample average^{12,13}. The presence of intrusive grains implies physical disturbance of the deposit, as also inferred from the pattern of ^{14}C -based ages shown by the 1993 set of samples and the elemental carbon sample at 142-cm depth. We have also calculated optical ages using a modified central-age model¹⁶ (G1 values in Table 2). The central-age estimates are likely to be too old, however, because of the influence of insufficiently bleached grains. The true age of fully bleached grains in each sample should be bracketed, therefore, by the central-age and minimum-age model estimates. The G1–G2 single-grain optical ages are in correct stratigraphic order; they bracket the ^{14}C -based ages obtained from elemental carbon fragments of $>125\text{ }\mu\text{m}$ at each level; and they constrain the

Table 2 Dose rates, palaeodoses and optical ages (and published thermoluminescence ages¹) for Jinmium*

Sample code (Pit square)	Depth (cm)	Radionuclide activities† (Bq kg ⁻¹)						Total dose rate‡ (Gy kyr ⁻¹)		Palaeodoses§ (Gy)	Optical age‡ (kyr)
		²³⁸ U	²²⁶ Ra	²¹⁰ Pb	²²⁸ Ra	²²⁸ Th	⁴⁰ K				
North face											
COOR 9/1 (C1/II)	34	12.2 ± 1.8	14.1 ± 0.4	14 ± 2	14.6 ± 0.7	15.0 ± 0.4	38 ± 3	0.80 ± 0.04	M	1.7 ± 0.3	2.1 ± 0.4
									S	1.9 ± 0.3	2.3 ± 0.3
									G1	1.0 ± 0.1	1.2 ± 0.2
									G2	0.43 ± 0.04	0.54 ± 0.06
									TL	2.4 ± 0.8	2.3 ± 0.7
COOR 1/3 (C1/III)	68	9.5 ± 0.9	15.4 ± 0.2	11.7 ± 1.0	18.1 ± 0.4	18.4 ± 0.3	38.3 ± 1.5	0.80 ± 0.04	M	3.5 ± 0.5	4.3 ± 0.6
									S	3.6 ± 0.2	4.5 ± 0.3
									G1	1.7 ± 0.3	2.2 ± 0.3
									G2	0.24 ± 0.03	0.30 ± 0.04
									TL	50 ± 6	50 ± 6
COOR 2/2 (C1/III)	110	8.1 ± 1.7	10.6 ± 0.3	13.8 ± 1.7	13.3 ± 0.7	12.6 ± 0.4	28 ± 5	0.73 ± 0.04	M	7.8 ± 0.8	10.7 ± 1.2
									S	8.2 ± 0.2	11.3 ± 0.7
									G1	5.0 ± 0.3	6.9 ± 0.6
									G2	1.4 ± 0.2	1.9 ± 0.2
									TL	63 ± 4	75 ± 7
South face											
COOR 8/1 (C1/III)	96	9.7 ± 1.1	9.4 ± 0.2	10.7 ± 1.2	11.3 ± 0.4	11.2 ± 0.2	16.2 ± 1.5	0.65 ± 0.03	M	6.0 ± 0.7	9.2 ± 1.1
									S	7.2 ± 0.4	11.1 ± 0.8
									G1	3.6 ± 0.2	5.6 ± 0.4
									G2	0.72 ± 0.05	1.11 ± 0.09
									TL	49 ± 5	58 ± 7
COOR 7/1 (C1/III)	115	9.5 ± 1.6	7.7 ± 0.3	12 ± 2	8.3 ± 0.6	9.2 ± 0.3	13 ± 3	0.60 ± 0.03	M	8.1 ± 1.3	14 ± 2
									S	11.1 ± 1.7	19 ± 3
									G1	12.4 ± 0.4	21.2 ± 1.3
									G2	4.1 ± 0.3	6.9 ± 0.6
									TL	1.5 ± 0.2	2.6 ± 0.3
COOR 6/2 (C1/III)	142	7.6 ± 1.2	8.3 ± 0.3	10.2 ± 1.6	9.3 ± 0.5	9.6 ± 0.2	13.6 ± 1.9	0.59 ± 0.03	M	74 ± 5	116 ± 12
									S	11.1 ± 1.7	19 ± 3
									G1	12.4 ± 0.4	21.2 ± 1.3
									G2	4.1 ± 0.3	6.9 ± 0.6
									TL	1.5 ± 0.2	2.6 ± 0.3
COOR 5/2 (C1/III)	158	9.5 ± 1.0	9.6 ± 0.2	12.2 ± 1.2	10.2 ± 0.4	11.3 ± 0.2	14.9 ± 1.5	0.65 ± 0.03	M	13 ± 2	19 ± 3
									S	16 ± 2	21 ± 3
									G1	16.8 ± 0.5	22.7 ± 1.2
									G2	6.6 ± 0.5	9.0 ± 0.8
									TL	1.9 ± 0.2	2.6 ± 0.3
COOR 4/1 (C1/III)	201	10.8 ± 1.3	9.7 ± 0.3	11.9 ± 1.3	12.4 ± 0.5	12.0 ± 0.3	25 ± 4	0.74 ± 0.04	M	145 ± 9	176 ± 16
									S	16 ± 2	21 ± 3
									G1	16.8 ± 0.5	22.7 ± 1.2
									G2	6.6 ± 0.5	9.0 ± 0.8
									TL	1.9 ± 0.2	2.6 ± 0.3

* Thermoluminescence data (in italics) are taken directly from ref. 1 for the nearest stratigraphic equivalents to the optical samples (see Fig. 1).

† Activities are converted to dose rates using published relationships^{27,28}. Concentrations of 1 p.p.m. of ^{238}U , 1 p.p.m. of ^{232}Th and 1% of ^{40}K correspond to activities of 12.4, 4.1 and 316 Bq kg⁻¹, respectively.

‡ Mean ± total (1σ) uncertainty, calculated as the quadratic sum of the random and systematic uncertainties. 1 kyr is 1,000 years.

§ Mean palaeodose ± standard error, obtained for multiple ($n = 48$) aliquots (M), single aliquots composed of multiple grains (S), and individual grains (G1 and G2). Single-aliquot palaeodoses include all 24 estimates obtained across the preheat plateau regions (160–300 °C). Single-grain palaeodoses are calculated using central-age (G1) and minimum-age (G2) models adapted from ref. 16. Values in parentheses are the number of grains used in the palaeodose determinations and the percentage of precise grains with palaeodoses of $>20\text{ Gy}$. Precise grains are those with relative errors of $\pm 20\%$ or better.

age of the basal layer to less than 10 kyr. Details of the statistical procedures and single-grain OSL results will be submitted to *Archaeometry*.

A Holocene age for the Jinmium deposit is consistent with the lack of stratigraphic evidence for long breaks in sedimentation and the lack of strong red pedogenic staining, a feature that typically occurs in the Pleistocene sand aprons near the rock shelter¹. The sedimentation rate of the sand aprons also concurs with that in the shelter if based on the optical and ¹⁴C-based ages, whereas the rate derived using the thermoluminescence ages is more than an order of magnitude lower¹. Evidence for insufficient bleaching of the thermoluminescence signal used for dating the Jinmium sediments is presented elsewhere, with interpretations that yield younger thermoluminescence ages^{17,18}. The latter ages, however, remain maximum estimates because of sample contamination by insufficiently bleached grains, as shown by optical dating of single grains. The single-grain approach also indicates some physical disturbance of the deposit, from the presence of a few grains with relatively young optical ages in the deeper layers. A search for such grains should be made at other sites as a check on stratigraphic integrity and on the translocation of particulate organic matter, which may affect ¹⁴C-based determinations on composite samples. The oldest (luminescence-based) ages for human colonization of Australia remain those of 50–60 kyr from Malakunanja II (ref. 19) and Nauwalabila I (ref. 20). These sites are currently under investigation using single-grain optical dating methods to test evidence^{17,19,20} that the sediments were bleached sufficiently before burial. □

Methods

Palaeodose determinations. Quartz grains of 90–125 µm in diameter were extracted from sediment samples and then etched in 40% HF acid for 45 min to remove the alpha-dosed outer rinds. Multiple-aliquot additive-dose palaeodoses were obtained using conventional quartz-dating procedures^{6,8}, namely a preheat of 220 °C for 5 min^{4,20}, optical stimulation (420–550 nm) at room temperature, and integration of OSL emissions (detected through two 3-mm U-340 filters) over the full period of illumination (250 s, using the last 25 s to define background). We applied a thermal-transfer correction²¹, which showed that thermal transfer is negligible in these samples. We obtained single-aliquot (multiple-grain) regenerative-dose palaeodoses using the same apparatus, 100-s stimulations at 125 °C, and the OSL signals arising from the natural dose (S_N), a subsequent test dose (T_N), a regenerative dose (S_R), and a second test dose (T_R)^{14,15}. Emissions were integrated over the first 20 s of illumination, using the final 20 s for background. The ratio of test dose (0.1 Gy) signals was used to correct for differences in OSL sensitivity between the S_N and S_R cycles, and the palaeodose was calculated as $(S_N/S_R) \times (T_R/T_N) \times$ regenerative dose. A preheat plateau test⁶ made on each sample (10-s preheats at 160–300 °C) always produced flat plateaux across the temperature range examined, indicating negligible thermal transfer^{14,15}. A 10-s preheat at 260 °C was used for all single-grain analyses, and preheat plateaux obtained for selected grains confirmed the validity of this preheat. Palaeodoses for single grains were determined from the OSL emitted in the first 1–5 s of illumination, using the formula given above and the measured T_R/T_N ratio (0.5 Gy test doses) or the weighted mean of all single-grain T_R/T_N ratios for the sample (the samples examined have mean ratios of 1.0–1.2), or a combination of these, to correct for any sensitivity changes. The weighting procedure improves palaeodose precisions for weakly luminescent grains, but our general conclusions are insensitive to the illumination time and T_R/T_N model chosen. An additional cycle of regenerative (S_{R2}) and test (T_{R2}) doses was given to some grains, the doses being identical to their earlier counterparts, to verify that the same correct dose (calculated as $(S_R/S_{R2}) \times (T_{R2}/T_R) \times$ regenerative dose) is calculated for grains that have received a known dose. This condition was satisfied for the four samples examined (240 grains in total), and supports the validity of the single-grain palaeodose estimates.

Dose-rate determinations. Gamma-ray dose rates were derived from high-resolution gamma-ray spectrometry analyses of powdered samples²² (which showed that a condition of secular equilibrium presently exists in the ²³⁸U and ²³²Th decay chains; see Table 2) and from field measurements using a portable

gamma-ray spectrometer. Cosmic-ray dose rates were estimated from published relationships²³, making allowance for site altitude, geomagnetic latitude, and thickness of rock overburden. Beta dose rates were calculated from high-resolution gamma-ray spectrometry analyses using an attenuation factor²⁴ of 0.93 ± 0.03 , and internal alpha dose rates were deduced from U and Th concentrations in aliquots of acid-etched quartz (measured by inductively-coupled plasma mass spectrometry) and an alpha efficiency 'a' value of 0.04 (ref. 25). Dose rates were calculated assuming a long-term mean water content of $1.5 \pm 0.5\%$ (field contents of 0.1–1.0% are apt to be below average, as samples were collected during the dry season). Ages increase by ~1% for each 1% increase in water content.

Received 21 January; accepted 16 March 1998.

- Fullagar, R. L. K., Price, D. M. & Head, L. M. Early human occupation of northern Australia: archaeology and thermoluminescence dating of Jinmium rock-shelter, Northern Territory. *Antiquity* **70**, 751–773 (1996).
- Bird, M. I. & Gröcke, D. R. Determination of the abundance and carbon isotope composition of elemental carbon in sediments. *Geochim. Cosmochim. Acta* **61**, 3413–3423 (1997).
- Tuniz, C. et al. Research and measurement program at the ANTARES AMS facility. *Nucl. Instrum. Methods Phys. Res. B* **123**, 73–78 (1997).
- David, B., Roberts, R., Tuniz, C., Jones, R. & Head, J. New optical and radiocarbon dates from Ngarrabullgan Cave, a Pleistocene archaeological site in Australia: implications for the comparability of time clocks and for the human colonization of Australia. *Antiquity* **71**, 183–188 (1997).
- Huntley, D. J., Godfrey-Smith, D. I. & Thewalt, M. L. W. Optical dating of sediments. *Nature* **313**, 105–107 (1985).
- Aitken, M. J. Optical dating: a non-specialist review. *Quat. Sci. Rev.* **13**, 503–508 (1994).
- Duller, G. A. T. Recent developments in luminescence dating of Quaternary sediments. *Prog. Phys. Geogr.* **20**, 127–145 (1996).
- Aitken, M. J. *An Introduction to Optical Dating* (Oxford Univ. Press, Oxford, 1998).
- Godfrey-Smith, D. I., Huntley, D. J. & Chen, W.-H. Optical dating studies of quartz and feldspar sediment extracts. *Quat. Sci. Rev.* **7**, 373–380 (1988).
- Lamothe, M., Balescu, S. & Auclair, M. Natural IRSL intensities and apparent luminescence ages of single feldspar grains extracted from partially bleached sediments. *Radiat. Meas.* **23**, 555–561 (1994).
- Lamothe, M. & Auclair, M. Assessing the datability of young sediments by IRSL using an intrinsic laboratory protocol. *Radiat. Meas.* **27**, 107–117 (1997).
- Roberts, R. et al. Luminescence dating of rock art and past environments using mud-wasp nests in northern Australia. *Nature* **387**, 696–699 (1997).
- Murray, A. S. & Roberts, R. G. Determining the burial time of single grains of quartz using optically stimulated luminescence. *Earth Planet. Sci. Lett.* **152**, 163–180 (1997).
- Murray, A. S. & Roberts, R. G. Measurement of the equivalent dose in quartz using a regenerative-dose single-aliquot protocol. *Radiat. Meas.* (in the press).
- Murray, A. S. & Mejdahl, V. Comparison of regenerative-dose single-aliquot and multiple-aliquot (SARA) protocols using heated quartz from archaeological sites. *Quat. Sci. Rev.* (submitted).
- Galbraith, R. F. & Laslett, G. M. Statistical models for mixed fission track ages. *Radiat. Meas.* **21**, 459–470 (1993).
- Roberts, R. G. Luminescence dating in archaeology: from origins to optical. *Radiat. Meas.* **27**, 819–892 (1997).
- Spooner, N. A. Human occupation at Jinmium, northern Australia: 116,000 years ago or much less? *Antiquity* **72**, 173–178 (1998).
- Roberts, R. G., Jones, R. & Smith, M. A. Thermoluminescence dating of a 50,000 year-old human occupation site in northern Australia. *Nature* **345**, 153–156 (1990).
- Roberts, R. G. et al. The human colonisation of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. *Quat. Sci. Rev.* **13**, 575–583 (1994).
- Huntley, D. J. & Clague, J. J. Optical dating of tsunami-laid sands. *Quat. Res.* **46**, 127–140 (1996).
- Murray, A. S., Marten, R., Johnston, A. & Martin, P. Analysis for naturally occurring radionuclides at environmental concentrations by gamma spectrometry. *J. Radioanal. Nucl. Chem. Articles* **115**, 263–288 (1987).
- Prescott, J. R. & Hutton, J. T. Cosmic ray contributions to dose rates for luminescence and ESR dating: large depths and long-term time variations. *Radiat. Meas.* **23**, 497–500 (1994).
- Mejdahl, V. Thermoluminescence dating: beta-dose attenuation in quartz grains. *Archaeometry* **21**, 61–72 (1979).
- Rees-Jones, J. & Tite, M. S. Optical dating results for British archaeological sediments. *Archaeometry* **39**, 177–187 (1997).
- Stuiver, M. & Reimer, P. J. Extended ¹⁴C data base and revised CALIB 3.0 ¹⁴C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
- Nambi, K. S. V. & Aitken, M. J. Annual dose conversion factors for TL and ESR dating. *Archaeometry* **28**, 202–205 (1986).
- Olley, J. M., Murray, A. & Roberts, R. G. The effects of disequilibria in the uranium and thorium decay chains on burial dose rates in fluvial sediments. *Quat. Sci. Rev.* **15**, 751–760 (1996).
- Galbraith, R. F. Graphical display of estimates having differing standard errors. *Technometrics* **30**, 271–281 (1988).
- Galbraith, R. F. The radial plot: graphical assessment of spread in ages. *Nucl. Tracks Radiat. Meas.* **17**, 207–214 (1990).

Acknowledgements. We thank the traditional owners and custodians (B. Simon, P. Carlton, others from Marralam, and the local Mirriuwung-Gajerrong people), the Waringarri Aboriginal Corporation, the Conservation Commission of the Northern Territory, the Aboriginal Areas Protection Authority, and the managers of Spirit Hills station for permission to undertake this study; and other members of the ANSTO AMS team (in particular D. Fink, M. Hotchkis and A. Smith) for their contribution, L. Frick and R. Maas for ICP-MS analyses, N. Minch for help with Fig. 1, D. Huntley for his dose-intercept program, and M. Aitken, D. Huntley, O. Lian, A. Murray and A. Wintle for comments. The Australian Institute of Aboriginal and Torres Strait Islander Studies, Australian National University and University of Wollongong funded the 1995 fieldwork, and the Australian Institute of Nuclear Science and Engineering funded the 1993 OZ sample AMS determinations. We acknowledge the support of the Australian Research Council through Queen Elizabeth II Fellowships to R.R. and M.B. and a Research Fellowship to R.F., and support from CSIRO Mathematical and Information Sciences (R.G.).

Correspondence and requests for materials should be addressed to R.R. (e-mail: bert.roberts@latrobe.edu.au).

Rapid evolution to terrestrial life in Jamaican crabs

Christoph D. Schubart*†, Rudolf Diesel*† & S. Blair Hedges‡

* Fakultät für Biologie I, VHF, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

‡ Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

† Present addresses: Department of Biology and Laboratory for Crustacean Research, University of Southwestern Louisiana, Lafayette, Louisiana 70504-2451, USA (C.D.S.); Max-Planck-Institut für Verhaltensphysiologie, Postfach 1564, 82305 Starnberg, Germany (R.D.)

Crabs of the family Grapsidae are abundant organisms in most intertidal communities. However, relatively few species live in complete independence of the sea¹. Of those species that do, Jamaica's nine endemic species of land crabs are unique in their exceptional adaptations to terrestrial life, which include the only active brood-care for larvae and juveniles known in crabs²⁻⁶. These adaptations, and the morphological similarity to a group of southeast Asian land-dwelling crabs, have raised the question of the number and age of land invasions of the Jamaican species. Here we present molecular evidence that Jamaican land crabs represent a single adaptive radiation from a marine ancestor that invaded terrestrial habitats only 4 million years (Myr) ago. A Late-Tertiary origin has also been found for lizards and frogs of Jamaica⁷⁻⁹ and probably reflects the Mid-Tertiary inundation of that island¹⁰.

To allow survival in the freshwater environment, all Jamaican crabs of the endemic genera *Sesarma* and *Metopaulias* undergo an abbreviated non-feeding larval development phase^{2,6,11}. Offspring survival is further enhanced by highly complex brood-care strategies in locally confined breeding habitats²⁻⁶. For example, the bromeliad crab, *M. depressus*, raises its young in water-filled bromeliad leaf axils. The mother crab manipulates water quality by removing detritus, circulating the water to oxygenate it, and carrying empty snail shells into leaf axils as both a calcium source and a pH buffer²⁻⁴.

She also protects the leaf axil against potential predators, including damselfly nymphs and spiders^{2,5}. The snail-shell crab, *S. jarvisi*, breeds in empty shells of the snail *Pleurodonte*; the crab either turns the snail shells upside down to collect rainwater or carries water into the shell⁶. Both species feed their offspring, which remain in the nursery habitat for several months^{2,6}. In the bromeliad crab, successive broods may coexist on the same host plant, resulting in the formation of family groups².

Species of the endemic Jamaican land crabs occupy various terrestrial and freshwater habitats and show different degrees of dependence on water. Morphological differences between these species are often pronounced and can be related to their habitats^{11,12}. For example, the flattened body of the bromeliad crab allows it to squeeze into the leaf axils of bromeliads. Such major changes in the body plan have made it difficult to study the phylogenetic history of these crabs by looking at morphological characters. Initially it was suggested that the radiation of land crabs arose from the only marine *Sesarma* species of Jamaica, *S. curacaoense*, or from a related stock^{11,13}. However, other morphological studies^{1,14-16} have suggested that at least some of the Jamaican land crabs are more closely related to freshwater species (genus *Sesarmoides*) from southeast Asia.

We obtained sequences from two mitochondrial genes, the large subunit ribosomal RNA (16S rRNA) and cytochrome oxidase I (COI), from 22 crab species of the family Grapsidae, including *Metopaulias* and all the American representatives of the genus *Sesarma*. We constructed phylogenetic trees using parsimony¹⁷ and distance methods¹⁸ (Fig. 1); these trees indicate that the Jamaican endemic terrestrial crabs are monophyletic, the result of a single colonization event. The closest relatives of the Jamaican terrestrial crabs seem to be marine intertidal American representatives of the genus *Sesarma* and not southeast Asian *Sesarmoides*. These results indicate that the bromeliad crab *M. depressus* should be placed in the same genus as the other Jamaican endemic crabs, of the genus *Sesarma*.

The well constrained dating of the geological closure of the Panama landbridge (3.1 Myr ago)^{19,20} allowed us to calibrate a molecular clock for the evolution of *Sesarma* using genetic distances between recognized trans-isthmian sister species groups (see Methods). We used this molecular clock to estimate a separation of the Jamaican endemic *Sesarma* and *Metopaulias* from their

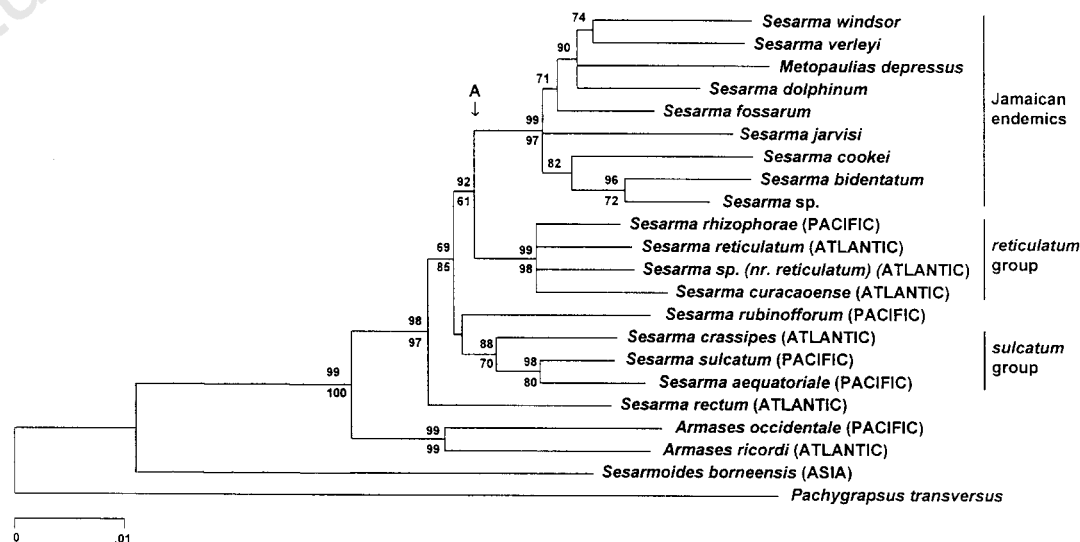


Figure 1 Molecular phylogeny of grapsid crabs, subfamily Sesarinae, inferred from an NJ analysis of DNA sequences of two mitochondrial genes, the large subunit (16S) rRNA and the cytochrome oxidase I genes. *Pachygrapsus transversus* (subfamily Grapsinae) was used to root the tree. Numbers are confidence values from an NJ analysis using the interior-branch method¹⁸ (above

lines) and an MP analysis using the bootstrap method (below lines). Interior branches of $d < 0.001$ and confidence values of $< 50\%$ not shown. The *reticulatum* and *sulcatum* groups contain the trans-isthmian species used in calibration of the molecular clock; node A represents the origin of the Jamaican lineage (see Methods). Scale bar, genetic distance.

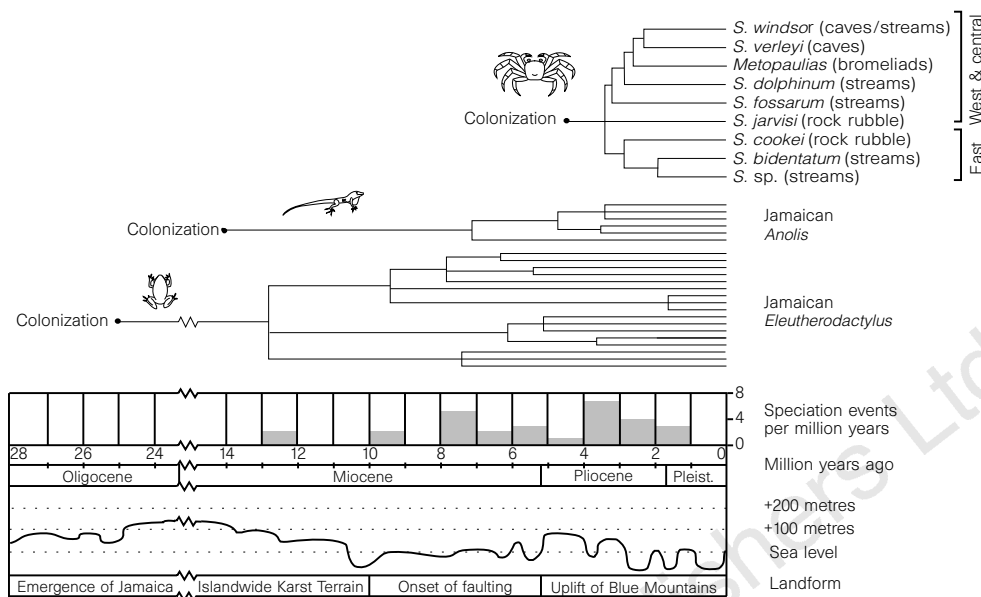


Figure 2 Phylogeny and adaptive radiation of the nine terrestrial crabs on Jamaica, with estimated times of divergence based on molecular clock calibrations. Phylogenetic trees of Jamaican lizards (*Anolis*)⁷ and frogs (*Eleutherodactylus*)⁸

are added for comparison of divergence times. The crab phylogeny is from Fig. 1. The histogram indicates the total number of speciation events per million years in all three groups. Sea-level changes are shown²⁶. Pleist., Pleistocene.

marine ancestor at 4.5 ± 0.42 Myr ago. Independent rate calibrations from two other studies of crustaceans^{21,22} also support a recent origin for the Jamaican lineage (see Methods). Invasion of freshwater and terrestrial habitats probably occurred between that time and the time of the earliest speciation events within Jamaica, at ~ 3.4 Myr ago (Fig. 2); subsequent speciation events occurred in the Late Pliocene epoch (3.4–1.9 Myr ago). Adaptive radiations of Jamaican lizards of the genus *Anolis*⁷ and frogs of the genus *Eleutherodactylus*⁸ also took place during the Late Miocene and Pliocene epochs (Fig. 2). These now are the best-documented terrestrial adaptive radiations of Jamaica and all three are relatively recent in origin. They occurred during the end of the Tertiary period, when terrestrial habitats became available for colonization after a Mid-Tertiary inundation of that island by the Caribbean Sea¹⁰.

The rapid radiation of Jamaican land crabs is characterized by an initial split into two geographic groups (Fig. 2), the position of *S. jarvisi* being unresolved. Both of these lineages include mountain-stream species and morphologically more derived and ecologically more specialized species. This suggests that two or more radiations took place simultaneously in distinct geographic regions of Jamaica, both independently giving rise to species with similar ecologies and morphologies. A similar geographic pattern is seen in the frogs of Jamaica⁸.

The colonization of terrestrial habitats by different animal groups is generally believed to be a long-term process involving many morphological, physiological and behavioural adaptations^{23,24}. Here it is shown that the complex adaptations of Jamaican terrestrial crabs^{2–6} evolved during a relatively short time span only a few million years ago. In contrast, marine species in the same genus separated for about the same amount of time (by the Isthmus of Panama) are ecologically and morphologically very similar²⁵. The Late-Tertiary emergence of Jamaica¹⁰ provided unoccupied terrestrial and freshwater habitats for exploitation by marine intertidal crabs and other organisms. The timing of speciation events in the subsequent radiations of Jamaican crabs, lizards and frogs coincides with unusually pronounced cycles of fluctuating sea levels²⁶, which may have isolated different populations (Fig. 2). Thus, the evolution of terrestrial habits in Jamaican crabs is another example of how

mechanisms of evolution can be better understood by the study of island life²⁷.

Methods

Amplification and sequencing. All crabs were collected by C.D.S. and R.D. in Jamaica, and North and Central America, and were preserved in 75% ethanol. Specimens of *Sesarma sulcatum* (Mexico), *Sesarma sp.* (nr. *reticulatum*) (Texas) and *Sesarmoides borneensis* (Borneo) were gifts. Genomic DNA was isolated from the muscle tissue of one of the crab's walking legs using a phenol/chloroform extraction. Selective amplification was carried out for portions of the two mitochondrial genes by polymerase chain reaction (PCR) (33–40 cycles: with 94°C/15 s, 50–55°C/15 s and 72°C/45 s denaturing, annealing and extension temperatures), using the primers 16sar (5'-CGCCTGTTTATCAA-AAACAT-3') and 16sbr (5'-ACGTGATCTGAGTTTCAGACCGG-3') (ref. 28), in addition to internal primers 5'-TGACCGTGCAAAGGTAGCATAA-3' and 5'-TTATCRCCCCAATAAAATA-3' (16L12 and 16H16, S.B.H. lab), for the 16S gene. For the COI gene, we used primers COIa (5'-AGTATAAGCGTCTGGGT AGTC-3'), COIf (5'-CCTGCAGGAGGAGGAGAYCC-3') (ref. 28) and the internal primers 5'-ATAATYTCYAYATYATTAAYCAAGA-3' and 5'-TTT GDGWTCRTGRARRGTWCTWARTCA-3') (COIL2 and COIH2, S.B.H. lab). Double-stranded PCR products were purified and used for asymmetric PCR (40–45 cycles: with 94°C/15 s, 55–60°C/15 s and 72°C/45 s). The resulting single-stranded DNA was filtered and sequenced by dideoxy chain termination with S35 radioactive labelling. The sequences have been deposited in the EMBL database (AJ225849–AJ225982).

Phylogenetic analysis. Alignments were done with ESEE²⁹. Of 1,073 total aligned sites, 345 were variable and 218 were informative for maximum parsimony (MP)¹⁷. Transversions were weighted 3 × transitions to correct for different substitution rates. Kimura 2-parameter distances were analysed by the neighbour-joining (NJ) method¹⁸. Statistical significance of resultant groups in the constructed trees was evaluated by the bootstrap method with 2,000 iterations (MP) or with an interior branch test (NJ)¹⁸.

Time estimation. Rate constancy³⁰ was rejected if all taxa were included, because of a faster rate of substitution in the Jamaican lineage, but was not rejected if that lineage was excluded. For this reason, divergence times were estimated using a lineage-specific method³⁰. In this method, the substitution rate was determined for non-Jamaican species and then used to estimate a date for the origin of the Jamaican lineage (node A in Fig. 1 and 'colonization' point in Fig. 2). This date, in turn, was used to estimate the substitution rate and

divergence times for the Jamaican lineage. To determine the non-Jamaican rate, we used the mean genetic distance ($d = 0.036$) of two *Sesarma* species group comparisons. In each comparison, the genetic distance was determined for species on either side of the Isthmus of Panama. For the *reticulatum* group, this was *S. rhizophorae* versus *S. reticulatum*, *S. curacaoense*, and *S. sp.* (nr. *reticulatum*), $d = 0.031$ (16S rRNA, 0.020; COI, 0.041). For the *sulcatum* group, this was *S. crassipes* versus *S. sulcatum* and *S. aequatoriale*; $d = 0.042$ (16S rRNA, 0.021; COI, 0.062). These trans-isthmian species of intertidal and supratidal crabs presumably have not been isolated for more than 3.1 Myr (refs 19, 20). The resulting average rate of pairwise sequence divergence (which equals the non-Jamaican rate) was 1.17% per Myr (0.65% for 16S rRNA and 1.66% for COI, if analysed separately).

The mean genetic distance between the Jamaican lineage and the *reticulatum* group is 0.063 ± 0.005 (s.e.m.). A relative rate test²⁸ revealed a distance of 0.037 on the Jamaican lineage and 0.026 on the *reticulatum* group lineage. The time of divergence between the two lineages (4.54 ± 0.42 Myr) was estimated on the *reticulatum* group lineage by applying the non-Jamaican rate; the standard error of the divergence times is the standard error of the mean genetic distance divided by this rate. That time estimate (4.54 Myr), in turn, was used to calibrate the rate of pairwise sequence divergence within the Jamaican lineage (1.63% per Myr; 0.88% for 16S rRNA and 2.33% for COI) and estimate divergence times among species (Fig. 2) from the length of internal branches in Fig. 1. A nearly identical time estimate (4.56 Myr) for the origin of the Jamaican lineage was obtained if a gamma correction¹⁸ ($\alpha = 1.5$) was used to account for rate variation among sites. In two other decapod crustacean studies, estimated rates of pairwise sequence divergence for these two genes were higher: 2.2% per Myr for 16S rRNA in hermit crabs²¹ and 2.2–2.6% per Myr for COI in snapping shrimps²². Although the rate calibrated within *Sesarma* is preferred, application of those independent rates would result in an even more recent time estimate (2.4–2.9 Myr ago) for the origin of the Jamaican lineage.

Received 10 November 1997; accepted 2 March 1998.

- Hartnoll, R. G. in *Biology of the Land Crabs* (eds Burggren, W. W. & McMahon, B. R.) 6–54 (Cambridge Univ. Press, Cambridge, 1988).
- Diesel, R. Parental care in an unusual environment: *Metopaulias depressus* (Decapoda: Grapsidae), a crab that lives in epiphytic bromeliads. *Anim. Behav.* **38**, 561–575 (1989).
- Diesel, R. Maternal control of calcium concentration in the larval nursery of the bromeliad crab, *Metopaulias depressus* (Grapsidae). *Proc. R. Soc. Lond. B* **264**, 1403–1406 (1997).
- Diesel, R. & Schuh, M. Maternal care in the bromeliad crab, *Metopaulias depressus* (Decapoda): maintaining oxygen, pH and calcium levels optimal for the larvae. *Behav. Ecol. Sociobiol.* **32**, 11–15 (1993).
- Diesel, R. Maternal care in the bromeliad crab, *Metopaulias depressus*: protection of larvae from predation by damselfly nymphs. *Anim. Behav.* **43**, 803–812 (1992).
- Diesel, R. & Horst, D. Breeding in a snail shell: ecology and biology of the Jamaican montane crab *Sesarma jarvisi* (Decapoda: Grapsidae). *J. Crust. Biol.* **15**, 179–195 (1995).
- Hedges, S. B. & Burnell, K. L. The Jamaican radiation of *Anolis* (Sauria: Iguanidae): an analysis of relationships and biogeography using sequential electrophoresis. *Carib. J. Sci.* **26**, 31–44 (1990).
- Hedges, S. B. An island radiation: allozyme evolution in Jamaican frogs of the genus *Eleutherodactylus* (Leptodactylidae). *Carib. J. Sci.* **25**, 123–147 (1989).
- Hass, C. A. & Hedges, S. B. Albumin evolution in West Indian frogs of the genus *Eleutherodactylus* (Leptodactylidae): Caribbean biogeography and a calibration of the albumin immunological clock. *J. Zool.* **225**, 413–426 (1991).
- Hedges, S. B. Historical biogeography of West Indian vertebrates. *Annu. Rev. Ecol. Syst.* **27**, 163–196 (1996).
- Hartnoll, R. G. The freshwater grapsid crabs of Jamaica. *Proc. Linn. Soc. Lond.* **175**, 145–169 (1964).
- Schubart, C. D., Reimer, J., Diesel, R. & Türkay, M. Taxonomy and ecology of two endemic freshwater crabs from western Jamaica with the description of a new *Sesarma* species (Brachyura: Grapsidae: Sesarmiinae). *J. Nat. Hist.* **31**, 403–419 (1997).
- Hartnoll, R. G. *Sesarma cookei* n. sp., a grapsid crab from Jamaica (Decapoda, Brachyura). *Crustaceana* **20**, 257–262 (1971).
- Serène, R. & Soh, C. L. New Indo-Pacific genera allied to *Sesarma* Say 1817 (Brachyura, Decapoda, Crustacea). *Treubia* **27**, 387–416 (1970).
- Holthuis, L. B. in *Stygofauna Mundi* (ed. Botosaneanu, L.) 589–615 (Leiden, 1986).
- Guinot, D. in *Encyclopedia Biogeologica* (eds Juberline, C. & Decou, V.) 165–179 (Société de Biopécologie, Moulis et Burest, 1994).
- Swofford, D. L. *Phylogenetic Analysis using Parsimony (PAUP)* (Univ. Illinois, Champaign, 1993).
- Kumar, S., Tamura, K. & Nei, M. MEGA: Molecular Evolutionary Genetics Analysis (Pennsylvania State Univ., University Park, 1993).
- Keigwin, L. Isotopic paleoceanography of the Caribbean and East Pacific: role of Panama uplift in late neogene time. *Science* **217**, 350–353 (1982).
- Coates, A. G. et al. Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. *Geol. Soc. Am. Bull.* **104**, 814–828 (1992).
- Cunningham, C. W., Blackstone, N. W. & Buss, L. W. Evolution of king crabs from hermit crabs ancestors. *Nature* **355**, 539–542 (1992).
- Knowlton, N., Weight, L. A., Solórzano, L. A., Mills, D. K. & Bermingham, E. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *Science* **260**, 1629–1632 (1993).
- Carroll, R. L. Revealing the patterns of macroevolution. *Nature* **381**, 19–20 (1996).
- Brusca, R. C. & Brusca, G. J. *Invertebrates* (Sinauer, Sunderland, MA, 1990).
- Abele, L. G. A review of the grapsid crab genus *Sesarma* (Crustacea: Decapoda: Grapsidae) in America, with the description of a new genus. *Smithson. Contrib. Zool.* **527**, 1–60 (1992).

- Haq, B. U., Hardenbol, J. & Vail, P. R. Chronology of fluctuating sea levels since the Triassic. *Science* **235**, 1156–1167 (1987).
- Williamson, M. *Island Populations* (Oxford Univ. Press, Oxford, 1981).
- Palumbi, S. R. et al. *The Simple Fool's Guide to PCR. A Collection of PCR Protocols* (University of Hawaii, Honolulu, 1991).
- Cabot, E. L. & Beckenbach, A. T. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Comp. Appl. Biosci.* **5**, 233–234 (1989).
- Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833 (1995).

Acknowledgements. We thank P. K. L. Ng, L. G. Abele and D. L. Felder for specimens; M. Schuh, G. Bäurle, D. Horst, J. A. Cuesta, J. Christy, DBML, STRI and IN.RE.NA.RE for support during collections; M. Frye, S. Kumar, and I. Ruvinsky for methodological advice; S. Kumar for the program PHYLTEST; E. Geißler and A. Beausang for help with figures; and C. A. Hass, S. Kumar, J. Reimer, and C. L. Schubart for comments on the manuscript. This project was funded by the DFG (C.D.S. and R.D.) and NSF (S.B.H.).

Correspondence and requests for materials should be addressed to C.D.S. (e-mail: cds5356@usl.edu).

The biosynthetic pathway of vitamin C in higher plants

Glen L. Wheeler, Mark A. Jones & Nicholas Smirnoff

School of Biological Sciences, University of Exeter, Hatherly Laboratories, Prince of Wales Road, Exeter EX4 4PS, UK

Vitamin C (L-ascorbic acid) has important antioxidant and metabolic functions in both plants and animals, but humans, and a few other animal species, have lost the capacity to synthesize it¹. Plant-derived ascorbate is thus the major source of vitamin C in the human diet. Although the biosynthetic pathway of L-ascorbic acid in animals is well understood², the plant pathway has remained unknown³—one of the few primary plant metabolic pathways for which this is the case. L-ascorbate is abundant in plants (found at concentrations of 1–5 mM in leaves and 25 mM in chloroplasts^{3,4}) and may have roles in photosynthesis and transmembrane electron transport^{3–5}. We found that D-mannose and L-galactose are efficient precursors for ascorbate synthesis and are interconverted by GDP-D-mannose-3,5-epimerase. We have identified an enzyme in pea and *Arabidopsis thaliana*, L-galactose dehydrogenase, that catalyses oxidation of L-galactose to L-galactono-1,4-lactone. We propose an ascorbate biosynthesis pathway involving GDP-D-mannose, GDP-L-galactose, L-galactose and L-galactono-1,4-lactone, and have synthesized ascorbate from GDP-D-mannose by way of these intermediates *in vitro*. The definition of this biosynthetic pathway should allow engineering of plants for increased ascorbate production, thus increasing their nutritional value and stress tolerance.

Evidence for the ascorbate-biosynthesis pathways so far proposed for plants is inconclusive. The most effective exogenous precursor of L-ascorbic acid is L-galactono-1,4-lactone, which is converted directly to ascorbate by a mitochondrial enzyme, L-galactono-1,4-lactone dehydrogenase^{6,7}. However, the role of L-galactono-1,4-lactone as a physiological precursor has been disputed⁸. L-galactono-1,4-lactone has not been detected in plants and, more significantly, its proposed production by D-galacturonic acid⁹ requires inversion of the hexose carbon skeleton, which extensive radiolabel tracer studies have shown does not occur during ascorbate synthesis from glucose¹⁰. An alternative pathway, involving the unusual intermediates D-glucosone and L-sorbose, has been suggested¹¹; this pathway involves no inversion of the carbon chain. However, the evidence that the physiological precursors are osones is also not conclusive.

If plants could produce L-galactono-1,4-lactone without inversion of the carbon chain, evidence for its involvement in ascorbate biosynthesis would be more powerful. To investigate the possible source of L-galactono-1,4-lactone in plants, we supplied L-galactose to barley leaf slices. This resulted in a rapid and substantial increase in the foliar ascorbate concentration, similar to that induced by L-

galactono-1,4-lactone (Fig. 1a). Feeding with L-galactose also increased ascorbate concentration in *A. thaliana* leaves and embryonic axes of germinating pea seedlings. The identity of the product as L-ascorbate was confirmed by its reactivity with ascorbate oxidase, by its ability to reduce acidic dichlorophenolindophenol (DCPIP), and by co-chromatography with L-ascorbate using thin-layer chromatography (TLC) and high-performance liquid chromatography (HPLC). Exogenous L-galactose accumulated in pea embryonic axes. The L-galactose was undetectable by gas chromatography (GC) 1 h after removal of the external supply, showing that it is rapidly and efficiently metabolized.

L-galactose is therefore an effective precursor of L-ascorbate in plants. Oxidation of L-galactose at position C1 would result in the production of L-galactono-1,4-lactone. An enzyme catalyzing this reaction was detected in cell-free extracts of *A. thaliana* leaves and pea embryonic axes. Both of these extracts supported L-galactose-dependent reduction of NAD⁺. The enzyme activity from pea embryonic axes was partially purified. The reduction activity with NADP⁺ was 10% of that with NAD⁺. The enzyme exhibited no

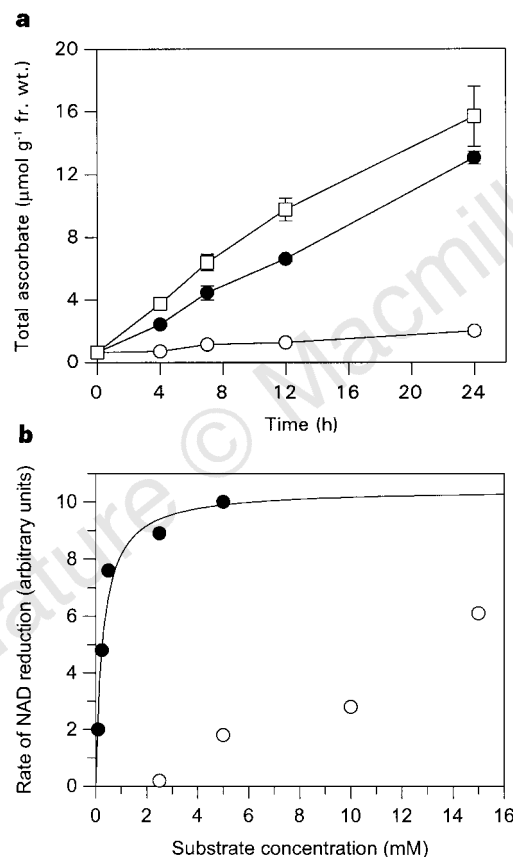


Figure 1 L-galactose is converted to L-ascorbate via L-galactono-1,4-lactone using L-galactose dehydrogenase. **a**, Ascorbate accumulation in barley leaves supplied with potential ascorbate precursors. Leaf segments were floated on water (open circles), 25 mM L-galactose (filled circles) or 25 mM L-galactono-1,4-lactone (open squares) in the light. The error bars indicate standard errors ($n = 3$). **b**, Identification of L-galactose dehydrogenase from pea embryonic axes. Activity was detected as L-galactose-dependent reduction of NAD⁺ and the reaction product was L-galactono-1,4-lactone (filled circles). The K_m for L-galactose was 0.3 mM and the pH optimum was pH 7.5–8.0. L-sorbose was oxidized with low affinity (open circles) and this oxidation activity co-purified with L-galactose dehydrogenase.

Table 1 Comparison of the metabolism of D-[U-¹⁴C]glucose and D-[U-¹⁴C]mannose by *A. thaliana* leaves

	¹⁴ C-glucose	¹⁴ C-mannose
Percentage of total ¹⁴ C		
Insoluble	50.1 ± 3.5	29.2 ± 2.6
Soluble	45.9 ± 2.9	67.9 ± 2.9
CO ₂	4.0 ± 1.0	2.9 ± 0.6
Ascorbic acid	1.0 ± 0.2	10.0 ± 1.2
Percentage of soluble ¹⁴ C		
Basic fraction	12.4 ± 5.8	4.1 ± 0.6
Neutral fraction	33.6 ± 5.8	62.7 ± 22.1
Acidic fraction		
60 mM formic acid eluent	11.2 ± 0.6	17.5 ± 2.3
2 M formic acid eluent	25.9 ± 5.1	19.5 ± 1.4

The leaves were fed labelled compounds through their petiole and incubated for 4 h in the light. ¹⁴CO₂ was trapped and the leaves extracted in 5% formic acid. The extract was neutralized and soluble components separated by ion-exchange chromatography into neutral, basic, and two acidic fractions these two acidic fractions were eluted with 60 mM and 2 M formic acid. Ascorbic acid elutes in the 60 mM fraction. Values are means ± standard error ($n = 3$).

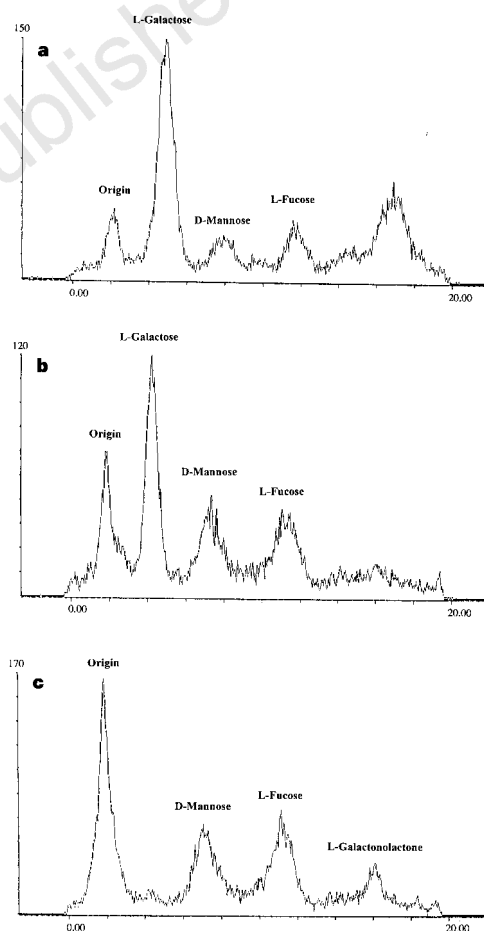


Figure 2 *In vitro* conversion of GDP-D-mannose to L-galactose and L-galactono-1,4-lactone by pea embryonic axis extracts. Crude extracts were incubated with GDP-D-[U-¹⁴C]mannose. After incubation, the reaction products were separated into neutral (free sugars) and acidic (phosphorylated or GDP-sugars) fractions by anion-exchange chromatography. The acidic fraction was then hydrolysed with trifluoroacetic acid to release free sugars. Sugars from both fractions were then separated by TLC and ¹⁴C label was detected by scanning the plates. **a**, TLC separation of sugars from acidic reaction products after 4 h incubation. Appearance of ¹⁴C in L-galactose shows the GDP-D-mannose was converted to GDP-L-galactose. GDP-L-fucose was also formed. **b**, Labelled L-galactose was detected in the neutral fraction of the reaction products, showing that free L-galactose was formed from GDP-D-mannose. **c**, Addition of 5 mM NAD⁺ for the final 3.5 h of the reaction shown in **b** resulted in disappearance of L-galactose and formation of L-galactono-1,4-lactone.

activity that was dependent on D-galactose, D-glucose, D-mannose, L-fucose and D-arabinose. The reaction product showed a positive response with the hydroxylamine assay for lactones, with 1:1 stoichiometry between NADH formation and lactone production (in L-galactono-1,4-lactone equivalents). We purified the reaction product and subjected it to GC-mass spectrometry (MS) to distinguish between two possible reaction products (C6 oxidation results in a uronic acid or lactone, whereas C1 oxidation results in an aldonic acid or lactone). We detected one peak, which co-chromatographed with authentic L-galactono-1,4-lactone. Its mass spectrum was identical to that of L-galactono-1,4-lactone. We assume that the initial product, from L-galactopyranose, was L-galactono-1,5-lactone which is relatively unstable and spontaneously converts to the 1,4 lactone. The K_m for L-galactose was 0.3 mM (Fig. 1b). The enzyme oxidized L-sorbose, another suggested intermediate in plant ascorbate biosynthesis¹¹, but with very low affinity (Fig. 1b). This activity had a similar preference for NAD⁺. The product would be L-ascorbate acid¹², explaining the incorporation of labelled L-sorbose into ascorbate¹¹. L-sorbose has the same configuration as L-galactose at C4 and C5 and the chiral centre at C3 is lost in ascorbate.

We have therefore demonstrated the existence of a new higher-plant enzyme, L-galactose dehydrogenase (L-galactose:NAD⁺ oxidoreductase), which forms L-galactono-1,4-lactone from L-galactose with high affinity and specificity. Other dehydrogenases that oxidize C1 of non-phosphorylated aldoses have been identified, for example D-galactose dehydrogenase in *Pseudomonas fluorescens*¹³ L-fucose dehydrogenase in mammals¹⁴ and fungi¹⁵ and D-arabinose dehydrogenase in *Candida albicans*¹⁶. The latter enzyme has been implicated in synthesis of D-arabino-1,4-lactone, which is the precursor of D-erythroascorbate, the yeast analogue of L-ascorbate¹⁶.

L-galactose constitutes a small proportion of the galactose in non-cellulosic cell-wall polymers^{17,18}. L-galactose residues in the cell wall are derived from GDP-L-galactose, which is formed by GDP-D-mannose-3,5-epimerase, a poorly characterized enzyme reported in *Chlorella* and flax seeds¹⁹⁻²¹. We have detected GDP-mannose-3,5-

epimerase activity in extracts of pea embryonic axes (Fig. 2a) and in ammonium-sulphate precipitates from *A. thaliana* leaves, using GDP-D-[U-¹⁴C]mannose (universally labelled GDP-D-mannose) as substrate. The products were free L-galactose and one or more acidic compounds, which produced L-galactose after hydrolysis. These acidic compounds must be GDP-L-galactose and/or L-galactose-1-phosphate (Fig. 2a, b). We confirmed the identity of the non-acidic product as L-galactose, rather than D-galactose, by its lack of reactivity with D-galactose dehydrogenase²². L-galactose is formed by hydrolysis of GDP-L-galactose; we do not yet know if L-galactose-1-phosphate is an intermediate. We suggest that formation of L-galactose resulted from a specific hydrolytic activity, because very little free mannose was released from GDP-mannose (Fig. 2b).

NAD⁺ added to the reaction mixture during the latter part of the incubation period resulted in loss of L-galactose from the neutral sugar fraction and appearance of radioactivity in a compound with the same mobility as L-galactono-1,4-lactone (Fig. 2c). Therefore, to confirm that L-galactono-1,4-lactone was produced from GDP-mannose, we prepared a pea embryonic axis extract containing intact mitochondria. Assays were carried out as above but with the addition of cytochrome *c* (1.5 mg ml⁻¹) as an electron acceptor for L-galactono-1,4-lactone dehydrogenase^{6,7}. We added unlabelled carrier ascorbate to the reaction mixture and then separated ascorbate by ion-exchange chromatography followed by TLC or HPLC. A radioactive peak that co-chromatographed with carrier L-ascorbate was detected in both systems. Treatment with ascorbate oxidase removed the radiolabelled HPLC peak. This clearly shows the ability of the tissue extract to produce ascorbate from GDP-D-mannose.

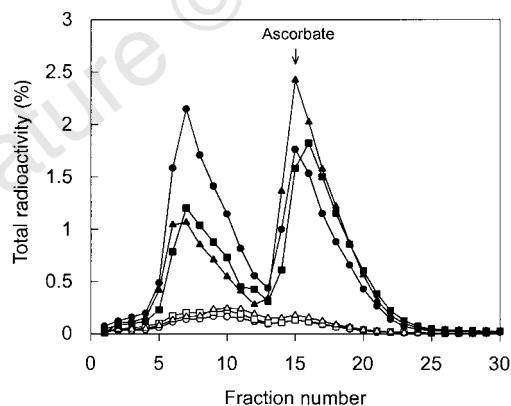


Figure 3 Comparison of the incorporation of D-[U-¹⁴C]glucose (open symbols) and D-[U-¹⁴C]mannose (closed symbols) into ascorbate by *A. thaliana* leaves. The leaves were fed labelled compounds through their petioles, incubated for 4 h in the light, and then extracted and fractionated as described in Table 1. The ¹⁴C content of ascorbic acid was determined after separation of the 60 mM formic acid fraction by HPLC, with correction for ascorbate loss during processing. The radioactive peak preceding ascorbate in the HPLC trace is largely dehydroascorbate.

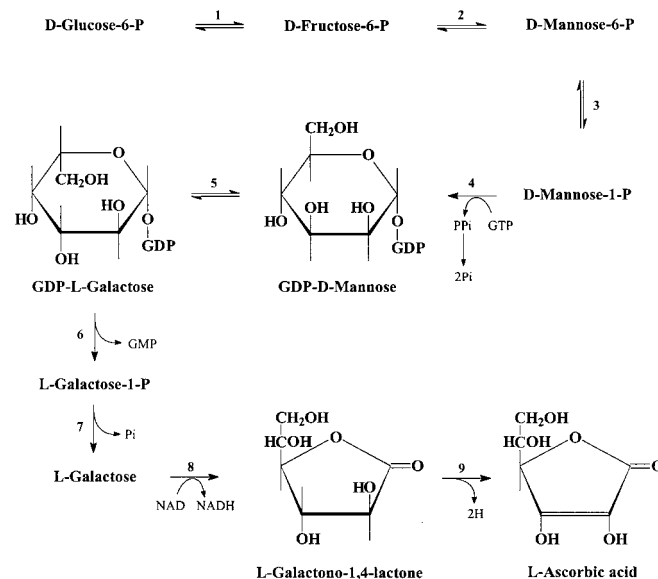


Figure 4 A proposed pathway for the biosynthesis of L-ascorbic acid in plants. The enzymes 1–5, which convert D-glucose-6-phosphate (D-glucose-6-P) to GDP-D-mannose and GDP-L-galactose, have been previously identified and provide GDP-sugar precursors for polysaccharide synthesis. L-galactose, the first dedicated intermediate, is provided by hydrolysis of GDP-L-galactose; a two-step hydrolysis (steps 6 and 7) through L-galactose-1-phosphate is shown, although this is speculative. L-galactose is oxidized at position C1 by L-galactose dehydrogenase (step 8, L-galactose:NAD⁺ oxidoreductase), forming L-galactono-1,4-lactone. This is oxidized by mitochondrial L-galactono-1,4-dehydrogenase (step 9) to L-ascorbic acid. Enzymes: 1, hexose phosphate isomerase; 2, phosphomannose isomerase; 3, phosphomannose mutase; 4, GDP-D-mannose pyrophosphorylase; 5, GDP-D-mannose-3,5-epimerase; 8, L-galactose dehydrogenase; 9, L-galactono-1,4-lactone dehydrogenase.

D-mannose is not extensively metabolized through glycolysis in plants and is instead mainly used in production of nucleotide sugar intermediates which are precursors for polysaccharides and glycoproteins¹⁷. Exogenous mannose is readily phosphorylated to mannose-6-phosphate^{17,23}. To confirm that the proposed biosynthetic pathway occurs *in vivo*, we fed D-[U-¹⁴C]mannose to *A. thaliana* leaves and compared its incorporation into ascorbate with incorporation of D-[U-¹⁴C]glucose (Fig. 3, Table 1). In agreement with earlier reports^{10,11,24}, 1% of the labelled glucose was converted to ascorbate. In contrast, 10% of the labelled mannose was incorporated into ascorbate, showing that this is a major metabolic fate for this sugar. The same result was found with marrow (*Cucurbita pepo*) roots. High concentrations of exogenous mannose did not affect ascorbate pool size, indicating that its conversion to L-galactose may be more rate-limiting than oxidation of L-galactose and L-galactono-1,4-lactone to produce ascorbate.

These results allow us to propose a pathway for ascorbate biosynthesis in higher plants that is consistent with earlier evidence (Fig. 4). We have demonstrated the ability of plant extracts to catalyse all the proposed reactions, including L-galactose oxidation by L-galactose dehydrogenase, a new enzyme in higher plants. The pathway reconciles the apparent contradiction between the ease of oxidation of L-galactono-1,4-lactone to ascorbate and earlier radiolabelling data. Pioneering work showed that, during conversion of D-glucose to ascorbate in plants, there is conservation of the hydroxymethyl group at C6, an epimerization at C5 and no inversion of the carbon chain^{8,10,11}. Our pathway, in which L-galactono-1,4-lactone is produced from D-glucose and D-mannose through GDP-D-mannose and L-galactose, is consistent with these requirements. We suggest that the ability of plants to incorporate radiolabel from the osones (D-glucosone and L-sorbose) into ascorbate also agrees with this scheme¹¹.

L-sorbose is a substrate analogue for L-galactose dehydrogenase (Fig. 1b). The very low affinity 'L-sorbose dehydrogenase' activity previously reported¹² could therefore be L-galactose dehydrogenase. However, this sorbose dehydrogenase activity, in contrast to L-galactose dehydrogenase, had a preference for NADP⁺ over NAD⁺ (ref. 12). The relationship between these activities requires investigation. More speculatively, if exogenous D-glucosone were reduced at position C2, about 50% of the product would be D-mannose which explains its incorporation into ascorbate¹¹.

Identification of the ascorbate-biosynthesis pathway fills a major gap in plant carbohydrate metabolism, as up to 10% of the soluble carbohydrate content of leaves can be L-ascorbate. We are now in a position to investigate the subcellular localization and control of ascorbate biosynthesis in plants and, ultimately, to manipulate its content with potential benefits for human nutrition and plant resistance to oxidative stress. An ascorbate-deficient *A. Thaliana* mutant (*vtc1*, formerly known as *soz1*) has been isolated²⁴; this mutant is affected in ascorbate biosynthesis²⁵. *vtc1*, and other biosynthesis mutants, will be valuable for testing the proposed pathway. □

Methods

Plant material. Primary leaves of 1-week-old barley (*Hordeum vulgare* cv Golden Promise) seedlings, *Arabidopsis thaliana* (Col-0) leaves and embryonic axes dissected from pea (*Pisum sativum* cv Meteor) seedlings 24 h after imbibition were used.

Identification of ascorbate as the *in vivo* product of L-galactose metabolism. Tissue was extracted, and ascorbate analysed, by ascorbate oxidase²⁵. Identity of ascorbate was confirmed by TLC combined with DCPIP staining²⁶ and HPLC²⁵.

L-galactose dehydrogenase activity. Tissue was homogenized in 50 mM Tris-HCl pH 7.5 containing 20% (v/v) glycerol, 1 mM EDTA and 2 mM dithiothreitol (DTT). Enzyme activity precipitating at between 50% and 75% ammonium sulphate saturation was further purified by hydrophobic interaction chromatography (phenyl sepharose). We assayed enzyme activity by

following NAD(P)H formation at 340 nm in 50 mM tris-HCl buffer, pH 7.5, with 0.1 mM NAD(P)⁺ and several concentrations of L-galactose. The identity of the reaction product was determined by a lactone assay¹⁶ and GC-MS. Before GC analysis, the product was delactonized by incubation for 2 h at pH 9.5 and was passed through a Dowex 1-formate column. After a water wash, the column was eluted with 0.1 M formic acid. The eluate was brought to 1 M formic acid and evaporated under an air stream to relactonize the product. The product was then dissolved in pyridine, derivatized with trimethylsilylimidazole with and without methoxyamine²⁷, and analysed by GC (Hewlett Packard Ultra 2 capillary column) and GC-MS (Hewlett Packard Ultra 1 capillary column coupled to a Hewlett Packard 5970 mass selective detector).

GDP-D-mannose-3,5-epimerase activity. Tissue (0.35 g fresh weight per ml) was homogenized in 100 mM tris-HCl buffer pH 7.6 containing 5 mM DTT, 1 mM EDTA and 1% polyvinylpyrrolidone before centrifugation at 12,000 g for 20 min. Extracts containing intact mitochondria were prepared by including 0.33 M sorbitol in the extraction medium as an osmoticum. The supernatant was then either used directly as a crude extract or precipitated with 90% saturation ammonium sulphate, and was then desalted with Sephadex G-25 (Pharmacia PD10 column). Crude extracts or precipitates (30 µl) were incubated with 3.7 kBq GDP-D-[U-¹⁴C]mannose (Amersham, UK) for measurement of ascorbate synthesis in 70 µl 25 mM Tris-HCl pH 8.0 containing 2 mM EDTA at 25 °C. The reaction products were passed through a SAX anion-exchange column (HPLC Technology, Macclesfield, UK). Neutral fractions were deionized with an SCX cation-exchange column (HPLC Technology, Macclesfield, UK) before TLC. GDP-sugars and sugar phosphates were eluted from the SAX column with 2 M formic acid and the free sugars were released by hydrolysis with 2 M trifluoroacetic acid at 80 °C for 1 h before TLC. Sugars were separated by TLC on silica plates (Whatman) impregnated with 0.3 M sodium dihydrogen orthophosphate, using an acetone/butanol/water (8:1:1 by volume) solvent²⁸. TLC plates were scanned with a Berthold Linear Analyser to detect radioactivity. Radioactive peaks were identified by co-chromatography with sugar standards detected by aniline/diphenylamine stain²⁹ and lactones by hydroxylamine/FeCl₃ (ref. 29). D- and L-galactose are not resolved, so the identity of the reaction product was confirmed by using D-galactose dehydrogenase to selectively remove this compound²².

[U-¹⁴C]glucose and [U-¹⁴C]mannose metabolism. Labelled compounds (Amersham, UK) were fed to cut leaves through the transpiration stream²⁵. The extracts were fractionated by ion-exchange chromatography. Ascorbate was further purified by HPLC and its ¹⁴C content determined²⁵.

Received 12 January; accepted 19 February 1998.

- Nishikimi, M., Fukuyama, R., Minoshima, S., Simizu, N. & Yagi, K. Cloning and chromosomal mapping of the human nonfunctional gene for L-gulonon-γ-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J. Biol. Chem.* **269**, 13685–13688 (1994).
- Burns, J. J. in *Metabolic Pathways* 2nd edn Vol. 1 (ed. Greenberg, D. M.) 394–411 (Academic, New York, 1967).
- Smirnoff, N. The function and metabolism of ascorbic acid in plants. *Ann. Bot.* **78**, 661–669 (1996).
- Foyer, C. H. in *Antioxidants in Higher Plants* (eds Alscher, R. G. & Hess, J. L.) 31–58 (CRC, Boca Raton, 1993).
- Horemans, N., Asard, H. & Caubergs, R. J. The role of ascorbate free radical as an electron acceptor to cytochrome *b*-mediated transmembrane electron transport in higher plants. *Plant Physiol.* **104**, 1455–1458 (1994).
- Mapson, L. W. & Breslow, E. Biological synthesis of L-ascorbic acid: L-galactono-γ-lactone dehydrogenase. *Biochem. J.* **68**, 395–406 (1958).
- Oba, K., Ishikawa, S., Nishikawa, M., Mizuno, H. & Yamamoto, T. Purification and properties of L-galactono-γ-lactone dehydrogenase, a key enzyme for ascorbic acid biosynthesis, from sweet potato roots. *J. Biochem.* **117**, 120–124 (1995).
- Loewus, F. A. & Loewus, M. B. Biosynthesis and metabolism of L-ascorbic acid in plants. *Crit. Rev. Plant Sci.* **5**, 101–119 (1987).
- Mapson, L. W. & Isherwood, F. A. Biological synthesis of L-ascorbic acid: the conversion of derivatives of D-galacturonic acid to L-ascorbate in plant extracts. *Biochem. J.* **64**, 13–22 (1956).
- Loewus, F. A. Tracer studies of ascorbic acid formation in plants. *Phytochemistry* **2**, 109–128 (1963).
- Saito, K., Nick, J. A. & Loewus, F. A. D-Glucosone and L-sorbose, putative intermediates of L-ascorbic acid biosynthesis in detected bean and spinach leaves. *Plant Physiol.* **94**, 1496–1500 (1990).
- Loewus, M. W., Bedgar, D. L., Saito, K. & Loewus, F. A. Conversion of L-sorbose to L-ascorbic acid by a NADP⁺-dependent dehydrogenase in bean and spinach leaf. *Plant Physiol.* **94**, 1492–1495 (1990).
- Maier, E. & Kurtz, G. D-galactose dehydrogenase from *Pseudomonas fluorescens*. *Methods Enzymol.* **89**, 176–181 (1982).
- Schachter, H., Sarney, J., McGuire, E. J. & Roseman, S. Isolation of diphosphopyridine nucleotide-dependent L-fucose dehydrogenase from pork liver. *J. Biol. Chem.* **244**, 4785–4792 (1969).
- Conter, P. F., Guimarães, M. F. & Veiga, L. A. Induction and repression of L-fucose dehydrogenase of *Pullularia pullulans*. *Can. J. Microbiol.* **30**, 753–757 (1984).
- Kim, S.-T., Huh, W.-K., Kim, J.-Y., Hwang, S.-W. & Kang, S.-O. D-Arabinose dehydrogenase and biosynthesis of erythroascorbate in *Candida albicans*. *Biochim. Biophys. Acta* **1297**, 1–8 (1996).
- Roberts, R. M. The metabolism of D-mannose-¹⁴C to polysaccharide in corn roots. Specific labelling of L-galactose, D-mannose, and L-fucose. *Arch. Biochem. Biophys.* **145**, 685–692 (1971).
- Baydoun, E. A.-H. & Fry, S. C. [³H]Mannose incorporation in cultured plant cells: investigation of L-galactose residues of the primary wall. *J. Plant Physiol.* **132**, 484–490 (1988).

19. Barber, G. A. Observations on the mechanisms of the reversible epimerization of GDP-D-mannose to GDP-L-galactose by an enzyme from *Chlorella pyrenoidosa*. *J. Biol. Chem.* **254**, 7600–7603 (1979).
20. Barber, G. A. Synthesis of L-galactose by plant enzyme systems. *Arch. Biochem. Biophys.* **147**, 619–623 (1971).
21. Feingold, D. S. In *Encyclopedia of Plant Physiology*, Vol. 13A (eds Loewus, F. A. & Tanner, W.) 3–76 (Springer, Berlin, 1982).
22. Roberts, R. M. & Harter, E. Determination of L-galactose in polysaccharide material. *Phytochemistry* **12**, 2679–2682 (1973).
23. Harris, G. C. *et al.* Mannose metabolism in corn and its impact on leaf metabolites, photosynthetic gas exchange, and chlorophyll fluorescence. *Plant Physiol.* **82**, 1081–1089 (1986).
24. Conklin, P. L., Williams, E. H. & Last, R. L. Environmental stress sensitivity of an ascorbic acid-deficient *Arabidopsis* mutant. *Proc. Natl Acad. Sci. USA* **93**, 9970–9974 (1996).
25. Conklin, P. L., Pallanca, J. E., Last, R. L. & Smirnoff, N. L-Ascorbic acid metabolism in the ascorbate-deficient *Arabidopsis* mutant *vtc1*. *Plant Physiol.* **115**, 1277–1285 (1997).
26. Chen, Y.-T., Isherwood, F. A. & Mapson, L. W. Quantitative estimation of ascorbic acid and related substances in biological extracts by separation on a paper chromatogram. *Biochem. J.* **55**, 821–823 (1953).
27. Andrews, M. A. Capillary gas-chromatographic analysis of monosaccharides: improvements and comparisons using trifluoroacetylation and trimethylsilylation of sugar O-benzyl and O-methyl oximes. *Carbohydrate Res.* **194**, 1–19 (1989).
28. Ghebregabher, M., Ruffini, S., Monaldi, B. & Lato, M. Thin layer chromatography of monosaccharides. *J. Chromatogr.* **127**, 133–162 (1976).
29. Dawson, R. M. C., Elliott, D. C., Elliott, W. H. & Jones, K. M. (eds) *Data for Biochemical Research* 2nd edn. (Clarendon, Oxford, 1969).

Acknowledgements. The research was supported by a BBSRC studentship (G.L.W.), a Nuffield Student Bursary (M.A.J.) and Zeneca Agrochemicals. We thank Nippon-Roche for the gift of L-sorbose. Earlier work by J. Pallanca, funded by the BBSRC BOMRIP programme, provided a basis for this research. We thank M. Raymond for technical assistance; J. Kingdon and J. Hindley for growing the plants; and W. Schuch for his support.

Correspondence and requests for materials should be addressed to N.S. (e-mail: N.Smirnoff@exeter.ac.uk).

Visual input evokes transient and strong shunting inhibition in visual cortical neurons

Lyle J. Borg-Graham, Cyril Monier & Yves Frégnac

Equipe Cognisciences, Institut Alfred Fessard, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France

The function and nature of inhibition of neurons in the visual cortex have been the focus of both experimental and theoretical investigations^{1–7}. There are two ways in which inhibition can suppress synaptic excitation^{2,8}. In hyperpolarizing inhibition, negative and positive currents sum linearly to produce a net change in membrane potential. In contrast, shunting inhibition acts nonlinearly by causing an increase in membrane conductance; this divides the amplitude of the excitatory response. Visually evoked changes in membrane conductance have been reported to be nonsignificant or weak, supporting the hyperpolarization mode of inhibition^{3,9–12}. Here we present a new approach to studying inhibition that is based on *in vivo* whole-cell voltage clamping. This technique allows the continuous measurement of conductance dynamics during visual activation. We show, in neurons of cat primary visual cortex, that the response to optimally orientated flashed bars can increase the somatic input conductance to more than three times that of the resting state. The short latency of the visually evoked peak of conductance, and its apparent reversal potential suggest a dominant contribution from γ -aminobutyric acid ((GABA)_A) receptor-mediated synapses. We propose that nonlinear shunting inhibition may act during the initial stage of visual cortical processing, setting the balance between opponent 'On' and 'Off' responses in different locations of the visual receptive field.

In the visual cortex, there are two major types of receptive fields, 'simple' and 'complex', based partly on the degree of spatial overlap between On and Off responses (evoked by an increase or decrease in light contrast, respectively) in the visual field^{4,13}. Simple receptive fields have distinct On and Off subfields, whereas these are over-

lapping in complex cells. We can consider how intracortical inhibition works during visual cortical processing at different levels. At the functional level, inhibition could help in the push-pull organization of opponent responses (for example, hyperpolarization is evoked by a decrease in light contrast in the On subfield) seen in simple receptive fields⁴. It could also control the spatial tuning of On and Off excitation, as intracortical blockade of γ -aminobutyric acid (GABA)_A receptors results in loss of segregation of On and Off excitatory responses in simple cells, whether measured extracellularly¹ or intracellularly¹⁴.

At the biophysical level, even if shunting inhibition does exist, there is still a quantitative issue concerning its functional importance. As the reversal potential for GABA_A-mediated channels, which are probably responsible for the shunt in membrane conductance, is near the resting potential, shunting inhibition must produce a large change in the postsynaptic conductance to significantly counteract excitation. Simulations show that this shunt should be visible as a 100–1000% increase in the somatic input conductance, $G_{in}(t)$, relative to the no-stimulus condition (characterized as G_{rest})^{2,15}. However, measurements from current clamp recordings have indicated limited conductance changes (relative

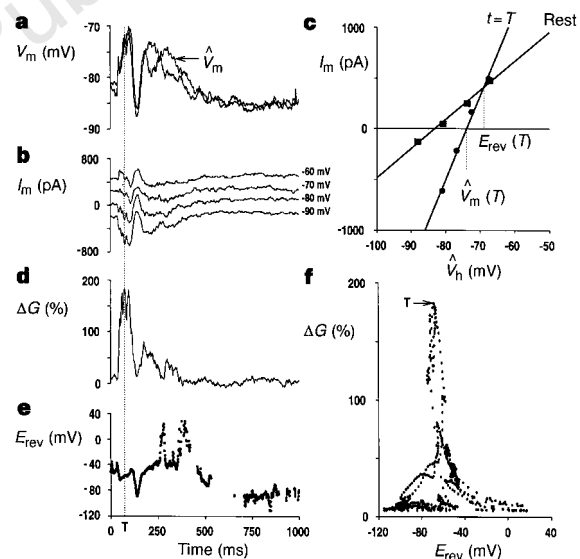


Figure 1 The visually evoked relative change in input conductance $\Delta G_{in}(t)$ and its apparent reversal potential $E_{rev}(t)$ are derived from the current waveforms measured by two to four voltage-clamp protocols, illustrated here for the subthreshold response of an end-stopped simple cell to an Off transition of a flashing bar (full response is shown in Fig. 2e, cell 4, position 8). **a**, Voltage responses under current clamp and predicted $\hat{V}_m$ (arrow) from voltage-clamp currents, assuming a linear model. **b**, Responses under voltage clamp for four command holding potentials. **c**, I/V characteristics derived from linear regressions corresponding to the resting state (squares), the slope of which gives G_{rest} , and during visual activation (circles) at the time T marked by a dotted line in (**a**–**e**), the slope of which gives $G_{in}(T)$. The voltage axis, $\hat{V}_h$, corresponds to the command holding potential corrected for R_s . The interpolated voltage at zero current of the I/V characteristic at any given time predicts the current clamp response ($\hat{V}_m(t)$) in **c**. The stability of the recording, and the justification of a linear approximation for this subthreshold example, are shown by the close match between the original and the interpolated traces superimposed in **a**. **d**, Relative $\Delta G_{in}(t)$, derived continuously over the complete duration of the visual stimulation. **e**, $E_{rev}(t)$. **f**, Phase plot of relative $\Delta G_{in}(t)$ versus $E_{rev}(t)$, where each point represents averages taken over 1 ms, illustrating the various trajectories in time (arrow for time T). In this case, the conductance response at time 0 is >105% of the input conductance at rest because the tail of the On response starts 1000 ms earlier. All responses in this paper are averages of 10 trials.

$\Delta G_{in}(t)$ typically in the order of 5–20%) during optimal and non-optimal visual activation^{3,8–11}.

We aimed to examine the balance between the excitation and inhibition underlying the spatial opponency of On and Off responses, and to identify the biophysical nature of the inhibition, using a new method for the continuous measurement of conductance dynamics. The standard approach to characterize $G_{in}(t)$ (used in the visual cortex *in vitro*^{16,17} and *in vivo*^{9,12,18}) is to measure voltage deflections in response to injected current pulses. A major constraint of this method is that although the rate of repetition of the pulses should be at least twice the highest frequency in the conductance signal, the pulse rate is limited by the resting time constant to account for the capacitive component of the input impedance. Voltage-dependent channels activated by the voltage change during a current pulse or by action potentials, and the electrical shunt introduced by sharp microelectrodes, also tend to underestimate the relative $\Delta G_{in}(t)$. Another approach¹⁰ relies on measuring the amplitude of electrically evoked test excitatory postsynaptic potentials (EPSPs) during the visual response. However, as the peak of the EPSP is reached within a few milliseconds, its amplitude approaches that of the peak response to a current impulse delivered at the soma, in this case given by the ratio of the EPSP driving force and cell capacitance. Thus, this technique is an insensitive measure of $\Delta G_{in}(t)$.

To overcome these constraints, we derived the entire $\Delta G_{in}(t)$ waveform from whole-cell, steady-state, voltage-clamp recordings. After characterization of receptive fields under current clamp (see, for example, Fig. 1a), we repeated the measurements of the response to the same visual stimulus under voltage clamp (continuous mode, no compensation for the access resistance, R_s) at two to four holding potentials, V_h (Fig. 1b). We then derived $G_{in}(t)$, where time is referenced to the visual stimulus, from the slope of the linear regression (I/V characteristic) of the points given by the measured averaged current, $I_m(t)$, and the holding potential, V_h , corrected for the drop across R_s ($\tilde{V}_h(t) = V_h(t) - R_s \times I_m(t)$) (Fig. 1c). The relative $\Delta G_{in}(t)$ was then taken as $100 \times (G_{in}(t) - G_{rest})/G_{rest}$ (%) (Fig. 1d).

As with any somatic-based impedance measurement, this method does not overcome the loss of visibility of dendritic inputs due to synaptic interactions or cable attenuation, or contamination by currents from poorly clamped voltage-dependent dendritic membrane^{19,20}. Rather, the advantage of using the voltage-clamp method is that distortion of synaptic events by transient voltage-dependent channels and capacitance near to the recording site are minimized. This method also allows the continuous estimation of the apparent somatic reversal potential of the visual response, $E_{rev}(t)$. This value is taken as the voltage of the intersection between the I/V characteristic at every time t and at rest (Fig. 1e) whenever the relative difference in their slopes (that is, relative $\Delta G_{in}(t)$) was

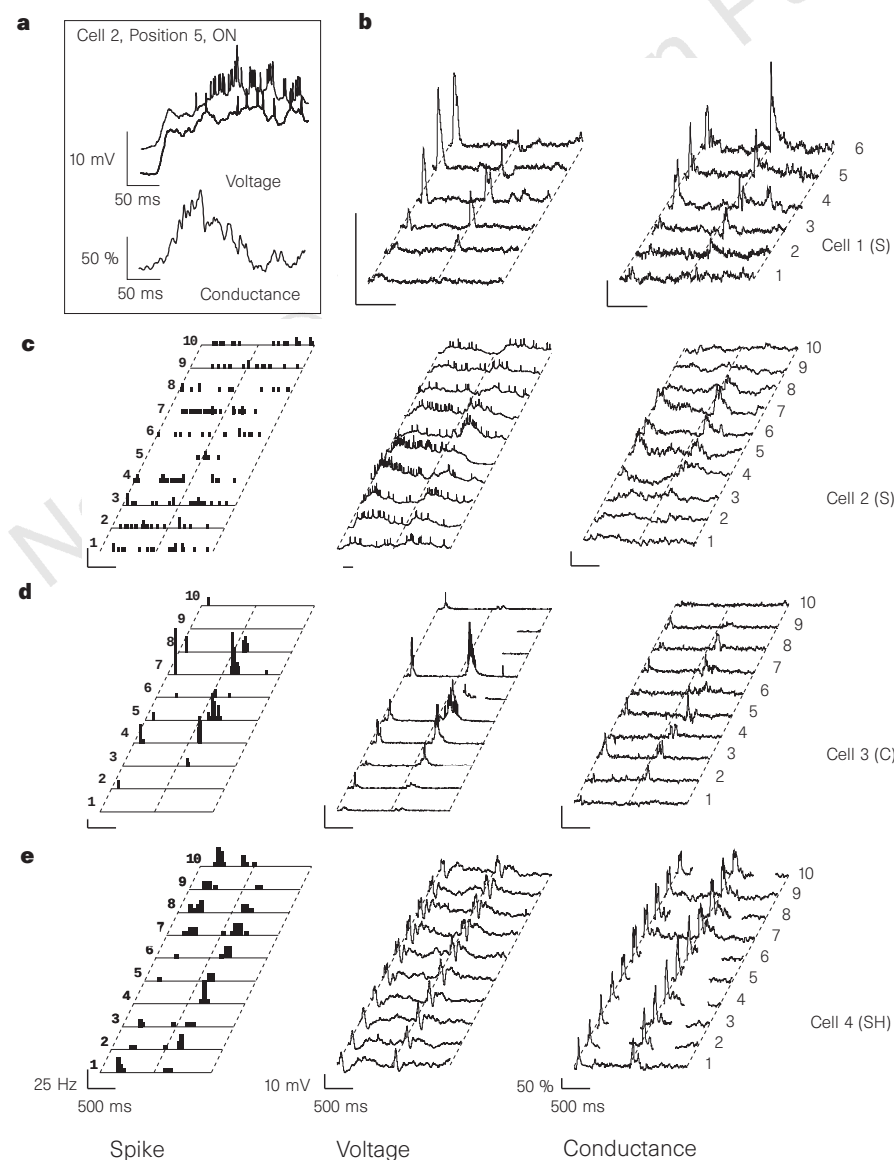


Figure 2 Response-plane receptive field maps based on spike activity (peri-stimulus histograms, PSTHs, left column), voltage (centre) and conductance (right) measurements. **a**, Time course of the voltage responses (top, 100 pA and bottom, 0 pA injected current) and of the conductance observed for the dominant On response of cell 2 in position 5. **b–e**, Response-plane maps for two simple cells (cells 1, 2, **b, c**), one complex cell (cell 3, **d**) and one end-stopped simple cell (cell 4, **e**). Stimuli are optimally oriented light bars flashed On and Off during two consecutive periods of 1 s, in adjacent locations in the visual field, spaced by 0.7° for cell 1, 0.5° for cell 2, and 1° for cells 3 and 4. For cell 4, a long bar stimulus was used to emphasize the inhibitory response. PSTHs are shown in the absence (**b, d**) or presence (**c, e**) of a constant depolarizing current (100 pA in **c**, 300 pA in **e**). G_{rest} and R_s values were, respectively, 17 nS and 60 M Ω for cell 1, 40 nS and 25 M Ω for cell 2, 15 nS and 20 M Ω for cell 3, and 29 nS and 15 M Ω for cell 4. Horizontal scale, 500 ms; vertical scales, 25 action potential per second for the PSTHs, 10 mV for the voltage traces, and 50% for relative $\Delta G_{in}(t)$.

>5%. This criterion allows a precise determination of the intersection point of the two characteristics. Assuming that the change in the visually evoked somatic conductance reflects the composite synaptic input reaching the soma, $E_{rev}(t)$ characterizes the balance between excitation and inhibition over time. Finally, to indicate more clearly the type of synaptic input underlying the conductance changes, we constructed phase plots of the relative $\Delta G_{in}(t)$ against $E_{rev}(t)$ (Fig. 1f).

The recordings from seven cells, from a total population of 109 cells first characterized in current clamp (see Methods), were stable enough in terms of R_s and visual responsiveness to allow reliable measurement of both visually evoked voltages and synaptic currents in receptive fields (response-plane maps²¹). Previous studies of On/Off opponency have concluded, on a qualitative basis, that the dominant and opponent responses in simple cells result from pure excitation and from pure subtractive inhibition, respectively^{4,9,22,23}. However, comparison of the temporal profiles of the post-stimulus time histograms (PSTHs), and the averaged evoked synaptic potential under current clamp, shows that both types of response in simple cells (Fig. 2b, c) may include composite excitatory/inhibitory input (for example, in Fig. 2c the dominant response in cell 2 position 5, and in Fig. 2b the opponent responses in cell 1, positions 5 and 6).

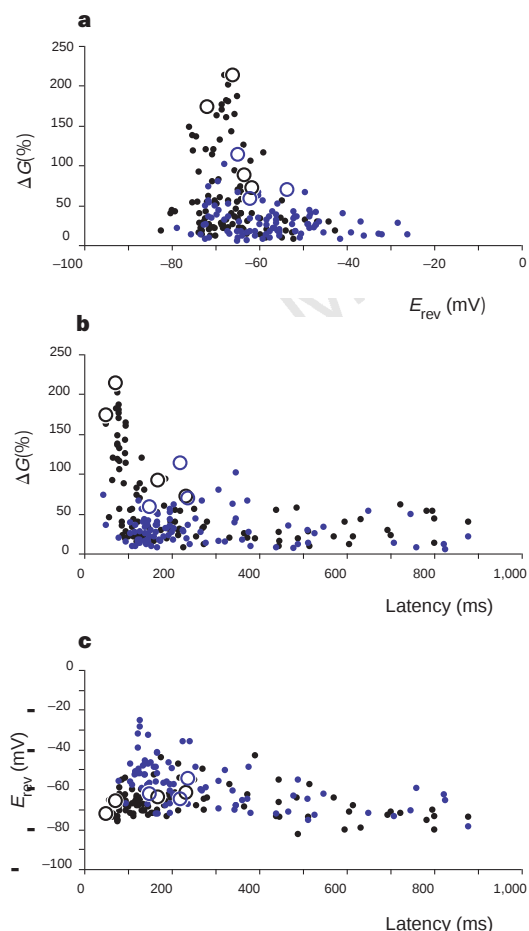


Figure 3 Relationships between amplitude, latency and apparent reversal potential of the maximum values of relative $\Delta G_{in}(t)$. Each point in the graphs corresponds to one stimulus condition (On or Off light or dark bar flashed at a given position). Results from seven cells (11 protocols) for all tested positions have been pooled. The responses with the maximum value of the relative $\Delta G_{in}(t)$ for each cell are marked by open circles. Black and blue symbols refer to simple and complex cells, respectively. **a**, Maximal ΔG versus E_{rev} . **b**, Maximal ΔG versus latency. **c**, E_{rev} versus latency.

The conductance profiles derived from the voltage-clamp measurements show that, for a given position in the receptive field, On and Off stimuli trigger a similar transient increase of $G_{in}(t)$ (Fig. 2). The average value for the largest relative conductance increases seen in each cell was 113% (s.d. = 58%, $n = 7$; see Fig. 3). In general, the maximum and global shape of the $\Delta G_{in}(t)$ waveform did not parallel the overall spike output or the waveform of the voltage record (see, for example, positions 5–6 for cell 1 in Fig. 2b, and positions 5–7 for cell 2 in Fig. 2c). Detailed examination shows that the rising phase and maximum of the conductance peaks are associated with an absence of spikes for both dominant and opponent responses, even when applying a depolarizing current to increase the cell's responsiveness (Fig. 2a, inset), indicating a transient early suppressive effect. The largest peaks ($138\% \pm 69\%$, $n = 4$) were observed for three simple cells and one end-stopped simple cell (Figs 2b, c, e and 4). Three complex cells showed smaller but still significant conductance transients ($80\% \pm 29\%$, $n = 3$; Figs 2d and 4).

Phase plots for relative $\Delta G_{in}(t)$ versus $E_{rev}(t)$ indicate a prototypical bounded region for the synaptic dynamics that is independent both of the resting potential of the cell and of its functional type (Fig. 4). The extrapolated E_{rev} peak values of all state trajectories converge near -60 to -70 mV. The averaged value of the actual peaks established for all seven cells ($-63.3 \text{ mV} \pm 5.5 \text{ mV}$), when compared to the reversal potentials of the major classes of currents gated by excitatory and inhibitory transmitters, concurs with the GABA_A reversal potential measured *in vitro* with identical patch pipette solutions ($-64 \pm 9 \text{ mV}$) (ref. 24). The main difference in the trajectories associated with specific On or Off responses is that, typically, more positive potentials were produced when the preferred stimulus was present (for example, the On response in position 6 for cell 1 and position 7 for cell 3; see Fig. 4), and more negative potentials were produced in response to both preferred and opponent stimuli (for example, cell 4 in Fig. 4e), indicating GABA_B receptor activation¹⁶. The trajectories often reached E_{rev} values beyond those expected for either GABA_B (less than -90 mV) or α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) ($>0 \text{ mV}$), which can be accounted for by voltage escape of distal inputs²⁰. The early onset of the larger conductance peaks, occurring sometimes as early as 50–70 ms after stimulation (Fig. 3b), further supports the conclusion that these responses are dominated by GABA_A rather than GABA_B activation (GABA_B activation occurs after longer latencies¹⁶). In some cases, early excitatory input is subsequently shunted by the large inhibitory input, with E_{rev} moving rapidly to a value close to the GABA_A reversal potential (Fig. 4a, chronogram inset). The consistent shape of the phase plots has two possible interpretations, not mutually exclusive. First, shunting inhibition is the largest component of synaptic activation that reaches the soma. Second, the shape of the peak may be constrained by the relative visibility at the soma of proximal GABA_A inputs compared with more distal excitatory (and perhaps GABA_B) inputs^{15,25}.

These findings are consistent with the two following observations. First, synaptic excitation and inhibition seem to overlap strongly during visual activation^{5,26} as is the case when synaptic activation is produced by electrical activation of thalamic or intracortical axons *in vitro*¹⁶ and *in vivo*¹⁸. Second, both electrical stimulation of intracortical pathways and iontophoretically applied GABA evoke a significant conductance increase *in vitro* as well as *in vivo*^{14,16,18}, contradicting the suggestion that, whatever the mode of activation, the effects of inhibition *in vivo* should differ from those obtained *in vitro*¹⁷.

These results argue against an extreme schematization of the synaptic basis for the spatial discreteness of On and Off subfields and opponency of their On and Off spike responses, according to which each subfield in simple cells is constructed exclusively from either On excitation only and Off inhibition only in On regions, or

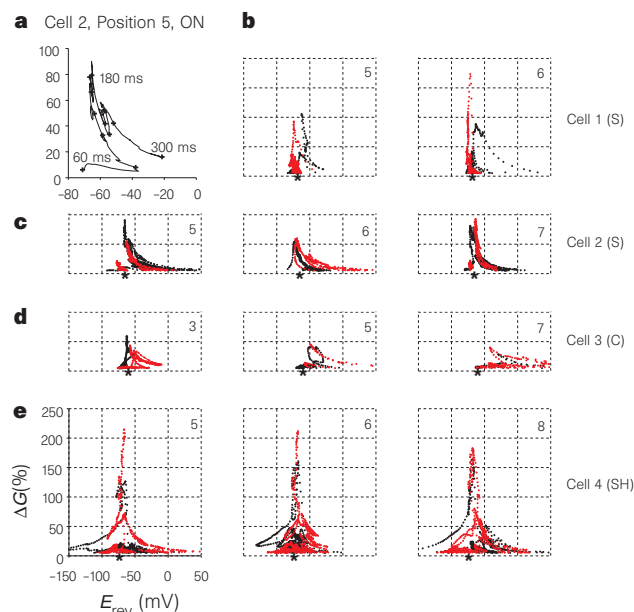


Figure 4 Phase plots of relative $\Delta G_{in}(t)$ versus $E_{rev}(t)$ for positions in the receptive fields (indicated in corner of each graph) eliciting the largest conductance or spike responses. **a**, The chronogram corresponds to the first 300 ms of the dominant response (0 nA) detailed in Fig. 2a. **b–e**, Plots correspond to the same cells shown in Fig. 2b–e. Asterisks under the plots indicate the cells' resting potentials (−76 mV for cell 1, −74 mV for cell 2, −68 mV for cell 3 and −80 mV for cell 4). Black and red dots correspond to the On and Off responses, respectively.

vice versa^{4,13}. Rather, the profiles of $E_{rev}(t)$ for dominant and opponent responses, and the study of synaptic responses under bicuculline application¹⁴, indicate that both On and Off responses normally combine excitatory and inhibitory inputs. We propose that excitatory and inhibitory synaptic inputs interact to shape the response for each On or Off transition of the stimulus by controlling the global gain of the cortical spiking response. A strongly shunting GABA_A input would be decisive during an early, nonlinear step of processing, allowing the dynamic emergence of the On/Off opponency expressed by small networks of reciprocally connected simple cells. If this shunting priming signal is weak, these circuits would then operate in a complex-like mode that lacks this form of functional selectivity. □

Methods

We recorded cells in the primary visual cortex of anaesthetized (Althesin), paralysed kittens and adult cats as described^{26,27}. For data-processing and visual-stimulation protocols we used in-house software. Patch electrode (3–5 MΩ) solutions contained 140 mM K-gluconate, 10 mM HEPES, 4 mM ATP, 2 mM MgCl₂, 0.4 mM GTP and 0.5 mM EGTA (KOH), with pH adjusted to 7.3 and the osmolarity adjusted to 285 mosM. We obtained whole-cell patch recordings (Axoclamp 2A amplifier) to depths $\leq 2,000 \mu\text{m}$. The estimate of R_s was revised as necessary over the course of the experiment and, in some cases, off-line by fitting the response to subthreshold hyperpolarizing current steps to the sum of two exponentials, with the faster exponential corresponding to the contribution of the electrode. Estimation of R_s was also confirmed by equalization of the height of the initial spike in response to different strengths of sustained depolarizing current. No significant correlation was found between R_s and the peak value of $G_{in}(t)$. We measured a tip offset potential of 10 mV, which was subtracted from the voltage records off-line²⁸.

We characterized electrophysiological and receptive-field properties under current clamp for 109 cells, each of which was recorded for >10 min (mean $\pm$ s.d. = 35 ± 26 min, range 11–146 min), seal resistance in attached mode $R_{seal} > 2 \text{ G}\Omega$ ($3.7 \pm 1.6 \text{ G}\Omega$, range 2–10 GΩ), and $R_s < 60 \text{ M}\Omega$ ($22 \pm 15 \text{ M}\Omega$, range 4–60 MΩ). In these cells the average input resistance R_{in} was $84.8 \text{ M}\Omega$ ($\pm 64.3 \text{ M}\Omega$, range 14–320 MΩ), the average membrane time constant τ_0 was 14.3 ms (± 12.4 ms), and the average resting potential was -72.8 mV ($\pm 9 \text{ mV}$).

We determined principal On and Off fields first by hand under current clamp using moving bars whose dimensions were adjusted to maximize firing, and then quantitatively by flashing optimally oriented bars (width 0.1–1.0 degrees) of both positive and negative contrasts at ten adjacent non-

overlapping positions across the receptive field. Visual-stimulation protocols were begun 10–15 minutes after whole-cell access, and lasted typically for 1 h. Receptive fields were classified as simple or complex on the basis of spike activity^{27,29}. In the case of reduced levels of visually evoked firing (see, for example, Fig. 2b), we classified a subfield as simple when there was a clear opponent hyperpolarization in the voltage record⁴.

Received 14 October 1997; accepted 2 March 1998.

- Sillito, A. M. The contribution of inhibitory mechanisms to the receptive field properties of neurones in striate cortex of the cat. *J. Physiol.* **250**, 305–329 (1975).
- Koch, C. & Poggio, T. In *Models of the Visual Cortex* (eds Rose, D. & Dobson, V. G.) 408–419 (J. Wiley and Sons, Chichester, 1985).
- Douglas, R. J., Martin, K. A. C. & Whitteridge, D. Selective responses of visual cortical cells do not depend on shunting inhibition. *Nature* **332**, 642–644 (1988).
- Ferster, D. Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J. Neurosci.* **8**, 1172–1180 (1988).
- Volgushev, M., Pei, X., Vidyasagar, T. R. & Creutzfeldt, O. D. Excitation and inhibition in orientation selectivity of cat visual cortex neurons revealed by whole-cell recordings *in vivo*. *Vis. Neurosci.* **10**, 1151–1155 (1993).
- Bush, P. C. & Sejnowski, T. J. Effects of inhibition and dendritic saturation in simulated neocortical pyramidal cells. *J. Neurophys.* **71**, 2183–2193 (1994).
- Carandini, M. & Heeger, D. J. Summation and division by neurons in primate visual cortex. *Science* **264**, 1333–1336 (1994).
- Eccles, J. C. *The Physiology of Synapses* (Springer, Berlin, 1964).
- Berman, N. J., Douglas, R. J., Martin, K. A. C. & Whitteridge, D. Mechanisms of inhibition in cat visual cortex. *J. Physiol.* **440**, 697–722 (1991).
- Ferster, D. & Jagadeesh, B. EPSP-IPSP interactions in cat visual cortex studied with *in vivo* whole-cell patch recording. *J. Neurosci.* **12**, 1262–1274 (1992).
- Carandini, M. & Ferster, D. A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science* **276**, 949–952 (1997).
- Pei, X., Volgushev, M., Vidyasagar, T. R. & Creutzfeldt, O. D. Whole-cell recording and conductance measurements in cat visual cortex *in vivo*. *Neuroreport* **2**, 485–488 (1991).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
- Shulz, D., Bringuier, V. & Frégnac, Y. Complex-like structure of simple visual cortical receptive fields is masked by GABA_A intracortical inhibition. *Soc. Neurosci. Abstr.* **19**, 638 (1993b).
- Koch, C., Douglas, R. & Wehmeier, U. Visibility of synaptically induced conductance changes: theory and simulations of anatomically characterized cortical pyramidal cells. *J. Neurosci.* **10**, 1728–1744 (1990).
- Connors, B. W., Malenka, R. C. & Silva, L. R. Two inhibitory postsynaptic potentials and GABA_A and GABA_B receptor mediated responses in neocortex of rat and cat. *J. Physiol.* **406**, 443–468 (1988).
- Berman, N. J., Douglas, R. J. & Martin, K. A. C. The conductances associated with inhibitory postsynaptic potentials are larger in visual cortical neurones *in vitro* than in similar neurones in intact, anaesthetized rats. *J. Physiol.* **418**, 107 (1989).
- Dreifuss, J. J., Kelly, J. S. & Krnjevic, K. Cortical inhibition and gamma-aminobutyric acid. *Exp. Brain Res.* **9**, 137–154 (1969).
- Rall, W. In *Neural Theory and Modeling* (ed. Reiss, R. F.) 73–97 (Stanford Univ., 1964).
- Spruston, N., Jaffe, D. B., Williams, S. H. & Johnston, D. Voltage- and space-clamp errors associated with the measurement of electronically remote synaptic events. *J. Neurophys.* **70**, 781–802 (1993).
- Palmer, L. A. & Davis, T. A. Receptive field structure in cat striate cortex. *J. Neurophys.* **46**, 260–295 (1981).
- Jagadeesh, B., Wheat, H. S. & Ferster, D. Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science* **262**, 1901–1904 (1993).
- Nelson, S., Toth, L., Sheth, B. & Sur, M. Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science* **265**, 774–777 (1994).
- Barker, J. L. & Harrison, N. L. Outward rectification of inhibitory postsynaptic currents in cultured rat hippocampal neurones. *J. Physiol.* **403**, 41–55 (1988).

25. Anderson, J. C., Douglas, R. J., Martin, K. A. C. & Nelson, J. C. Map of the synapses formed with the dendrites of spiny stellate neurons of cat visual cortex. *J. Comp. Neurol.* **341**, 25–38 (1994).
26. Bringuier, V., Frégnac, Y., Baranyi, A., Debanne, D. & Shulz, D. Synaptic origin and stimulus dependency of neuronal oscillatory activity in the primary visual cortex of the cat. *J. Physiol.* **500**, 751–774 (1997).
27. Debanne, D., Shulz, D. & Frégnac, Y. Activity-dependent regulation of ON and OFF responses in cat visual cortical receptive fields. *J. Physiol.* **500**, 523–548 (1998).
28. Neher, E. in *Methods in Enzymology: Ion Channels* (eds Rudy, B. & Iverson, L. E.) 123–131 (Academic, New York, 1992).
29. Orban, G. A. *Neuronal Operations in the Visual Cortex* (Springer, Berlin, Heidelberg, 1984).

Acknowledgements. We thank V. Bringuier and F. Chavane for help with some experiments; N. Gazeris, T. Bal, K. Grant, R. Kado, P.-M. Lledo, N. Ropert and D. Shulz for comments; and G. Sadoc and L. Glaeser for software assistance. This work was supported by HFSP and GIS Cognisciences grants (to Y.F.). L.J.B.G. was funded by fellowships from the CNRS and Foundations Philippe and Fyssen.

Correspondence and requests for materials should be addressed to Y.F. (e-mail: fregnac@iaf.cnrs-gif.fr).

Acute stress facilitates long-lasting changes in cholinergic gene expression

Daniela Kaufer*, Alon Friedman*†, Shlomo Seidman* & Hermona Soreq*

* Department of Biological Chemistry, The Alexander Silberman Life Sciences Institute, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

† Departments of Physiology and Neurosurgery, Faculty of Health Science, and Zlotowski Center for Neuroscience, Ben Gurion University, Beersheva 84105, Israel

Acute traumatic stress may lead to post-traumatic stress disorder (PTSD)¹, which is characterized by delayed neuropsychiatric symptoms including depression, irritability, and impaired cognitive performance². Curiously, inhibitors of the acetylcholine-hydrolysing enzyme acetylcholinesterase may induce psychopathologies that are reminiscent of PTSD^{3,4}. It is unknown how a single stressful event mediates long-term neuronal plasticity. Moreover, no mechanism has been proposed to explain the convergent neuropsychological outcomes of stress and of acetylcholinesterase inhibition. However, acute stress elicits a transient increase in the amounts released of the neurotransmitter acetylcholine and a phase of enhanced neuronal excitability⁵. Inhibitors of acetylcholinesterase also promote enhanced electrical brain activity⁶, presumably by increasing the survival of acetylcholine at the synapse. Here we report that there is similar bidirectional modulation of genes that regulate acetylcholine availability after stress and blockade of acetylcholinesterase. These calcium-dependent changes in gene expression coincide with phases of rapid enhancement and delayed depression of neuronal excitability. Both of these phases are mediated by muscarinic acetylcholine receptors. Our results suggest a model in which robust cholinergic stimulation triggers rapid induction of the gene encoding the transcription factor c-Fos. This protein then mediates selective regulatory effects on the long-lasting activities of genes involved in acetylcholine metabolism.

The molecular mechanisms translating a traumatic life experience into long-term neuropsychological sequelae are expected to involve complex changes in gene regulation. We have previously shown that adult FVB/N mice subjected to either forced swimming stress or inhibitors of the acetylcholine-hydrolysing enzyme acetylcholinesterase (AChE) exhibit dramatic increases in levels of messenger RNA encoding the early immediate transcription factor c-Fos in the brain⁷. *In vitro*, sagittal corticohippocampal brain slices exposed to AChE inhibitors showed enhanced neuronal excitability and similar increases in cortical c-fos gene expression within 10 min (Fig. 1a). These increases are mediated by cholinergic stimulation of muscarinic acetylcholine receptors.

The presence of c-Fos-binding sites in the promoters of key cholinergic genes, such as the genes encoding AChE^{8,9}, the acetyl-

choline-synthesizing enzyme choline acetyltransferase (ChAT)¹⁰, and the vesicular acetylcholine transporter (VACHT)¹¹, indicated that elevated c-Fos levels might activate regulatory pathway(s) leading to long-term changes in the expression of proteins mediating brain cholinergic neurotransmission. We performed quantitative reverse transcription with polymerase chain reaction (RT-PCR) on cortical RNA extracted either from mice 10–90 min after forced swimming or from brain slices after exposure to the cholinesterase

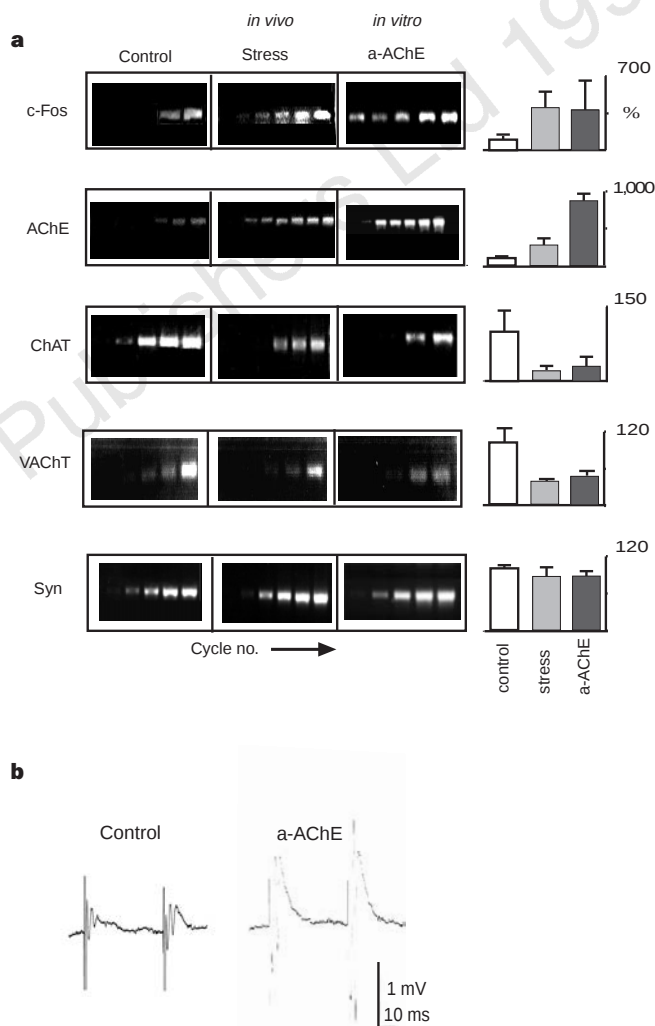


Figure 1 Acute stress and anticholinesterases modulate CNS gene expression similarly. **a**, RT-PCR analysis was performed on RNA extracted from the cortex of control mice and stressed mice or from sagittal corticohippocampal slices incubated with 1 μ M DFP or 1 mM pyridostigmine (a-AChE). Products were sampled every third cycle, electrophoresed, and stained with ethidium bromide. Data reflect c-Fos mRNA levels 10 min after stress or AChE inhibition, and AChE, ChAT, VACHT and synaptophysin (Syn) RNA levels 30 min after treatment. The figure shows representative gels and relative band intensities (mean $\pm$ s.d.), calculated from densitometric analysis of a single cycle verified to be within the linear range of product accumulation for the specific PCR reaction. On average, five RNA samples were analysed for each treatment group. For c-Fos, AChE, ChAT and VACHT, the differences in RNA levels observed between the control and either stress or a-AChE samples were all found to be statistically significant ($P < 0.02$) in a two-tailed Student's *t*-test. RNA from non-treated control animals generated patterns similar to those from non-treated slices (not shown). **b**, AChE inhibition increases neuronal excitability. The figure shows extracellular evoked potentials recorded in the CA1 area before (control) or 30 min following (a-AChE) addition of 1 μ M DFP to the perfusing solution. One of five reproducible experiments.

inhibitors diisopropylfluorophosphonate (DFP) or pyridostigmine. In all cases, AChE mRNA levels were markedly increased, whereas the levels of ChAT and VAcT mRNAs were reduced (Fig. 1a). These changes in gene expression lagged behind the increase in c-Fos levels by 20 min (Fig. 1a, and data not shown), consistent with the idea that c-Fos may activate or suppress cholinergic gene expression. The levels of mRNAs encoding synaptophysin, the L-type Ca^{2+} channel or glyceraldehyde phosphodehydrogenase remained unchanged (Fig. 1a, and data not shown). These results indicate that acute cholinergic stimulation promotes selective bidirectional changes in the expression of genes regulating acetylcholine metabolism. The combined effects of these changes should reduce the bioavailability of acetylcholine through suppressed synthesis/packaging and enhanced hydrolysis. It has been reported that transient increases and delayed reductions in acetylcholine levels accompany stress⁵. As modulated gene expression occurs *in vitro*, independently of the pituitary–adrenocortical axis, local mechanisms in the central nervous system are apparently enough to mediate this response.

Delayed reduction in acetylcholine levels after c-Fos induction predicts a secondary phase of suppressed neuronal excitability following both stress and AChE inhibition. To determine whether acetylcholinesterase inhibitors mediate both acute and delayed phases of cholinergic activity, we recorded extracellular potentials in the cell-body layer of the CA1 region of hippocampal slices; these potentials were evoked by orthodromic stimulation of the CA2/CA3 region of the stratum oriens enriched with cholinergic fibers¹². Exposure to various AChE inhibitors prompted, within 1 h, increased population spike amplitude, rate of rise, and duration

of paired-pulse facilitation under several stimulus intensities (Fig. 1b). When we extended AChE inhibition to 3 h, the augmented synaptic response and the population spikes were significantly muted, approaching responses seen under control conditions (Fig. 2a, b). These results show that AChE inhibitors mediate a transient, early phase of enhanced excitability that is followed by a delayed phase of suppressed neuronal activity. The non-hydrolysable acetylcholine analogue carbamylcholine promoted a similar and dose-dependent increase in amplitude and rate of rise of evoked population spikes (Fig. 2c). We could reversibly block both phases of inhibitor-mediated responses by adding the muscarinic antagonist atropine to the perfusion solution during the early phase (Fig. 2d). This indicates that the late phase of depressed activity represents a delayed response to a previous phase of acute muscarinic stimulation.

The c-fos gene includes a Ca^{2+} -responsive element and elevated c-Fos levels are a marker of neuronal hyperexcitation¹³. We therefore proposed that neuronal activity and/or intracellular accumulation of calcium play a role in translating the transient phase of cholinergic hyperactivation into changes in gene expression. The intracellular calcium chelator 1,2-bis-(2-aminophenoxy)ethane-*N,N,N,N'*-tetraacetic acid tetra(acetoxymethyl)ester (BAPTA-AM) prevented enhanced paired-pulse facilitation in response to physostigmine (Fig. 3a), and both BAPTA-AM and the sodium-channel blocker tetrodotoxin (TTX) attenuated the changes in c-Fos and ChAT mRNA levels that are mediated by AChE inhibition (Fig. 3b). Our results indicate that intense cholinergic activation initiates a calcium- and neuronal-activity-dependent feedback

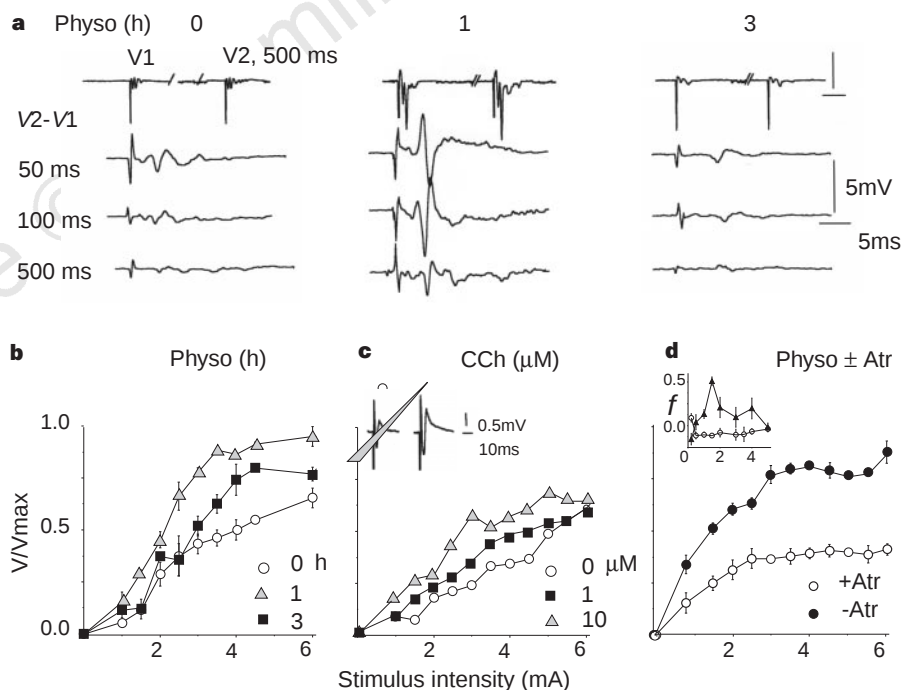


Figure 2. Delayed suppression of the hyperexcitation evoked by AChE inhibition. **a**, Enhancement of paired-pulse facilitation is transient. The figure shows the first and second evoked potentials (V_1 , V_2) separated by a 500-ms interval or the difference ($V_2 - V_1$) at intervals 50, 100 or 500 ms for hippocampal slices under control conditions (0 h) or following 1 or 3 h of perfusion with the carbamate AChE inhibitor physostigmine (physo, 10 μM). **b**, Delayed repression of increased population spike amplitudes. Average ($\pm$ s.d.) relative amplitudes (V/V_{max}) of evoked population spikes under control conditions (open circles), or following 1 or 3 h of continuous perfusion of 10 μM physostigmine (physo) (grey triangles or filled squares, respectively), are shown. V_{max} for control is 1.04 mV at 9.5 mA. **c**, Carbachol increases population spike amplitudes. The electrophysiology is as in

b, Control, empty circles; in presence of 1 μM carbamylcholine (CCh), filled squares; 10 μM CCh, grey triangles. V_{max} for control is 1.98 mV at 9.5 mA. Inset shows the average of ten responses in control solution (left) or following 30-min perfusion with 10 μM CCh (right). **d**, Increased population spikes are mediated by muscarinic receptor activation. Results are shown from 1 h after coperfusion of physostigmine (10 μM) and atropine (Atr, 1 μM) (open circles) and 2 h later, following atropine washout (filled circles). The inset shows values of paired-pulse facilitation ($f = (V_2 - V_1)/V_1$) in the presence of physostigmine alone (triangles) as compared with physostigmine + atropine (circles). V_{max} for control is 2.25 mV at 7.5 mA.

pathway that works to suppress cholinergic neurotransmission through modulation of protein synthesis.

AChE activity in the neocortex and hippocampus, but not cerebellum, of animals exposed to a single stress session increased by two- to threefold within 50 min after stress and cortical activity remained significantly higher than that in control mice for over 80 h (Fig. 4; $P < 0.05$, two-tailed Student's *t*-test). The physiological relevance of this elevated brain AChE activity was seen as resistance to an anticholinesterase-induced drop in body temperature following stress. Intraperitoneal injection of pyridostigmine (0.2 mg kg^{-1}) induced a $0.9 \pm 0.3^\circ\text{C}$ reduction in rectal temperature within 30 min in control mice ($n = 5$); no such drop was seen in mice subjected to forced swimming 24 h before injection of pyridostigmine ($0.0 \pm 0.3^\circ\text{C}$, $n = 5$, $P < 0.005$, two-tailed Student's *t*-test). This observation indicates occlusion of the inhibitor effect by the effects of stress, and strengthens the contention that stress and cholinergic activation act on the brain through a common pathway involving elevated levels of synaptic acetylcholine.

Non-denaturing polyacrylamide gel electrophoresis revealed new, quickly migrating AChE form(s) in the brains of stressed mice (Fig. 4, inset). The pattern of gel migration corresponded closely with that of secreted, monomeric recombinant read-through AChE produced in *Xenopus* oocytes (Fig. 4 inset)¹⁴. Although a minor mRNA species encoding read-through AChE was previously detected in brain and in several tumour cell lines¹⁵, its protein product has never been unequivocally identified *in vivo*. Following both stress and exposure to AChE inhibitors, a pronounced increase was observed in levels of this unspliced mRNA species in which pseudo-intron 4 is retained in the mature transcript (Fig. 5a). In contrast, no changes were seen in either the transcript containing the alternative 3' exon 6 and encoding the dominant synaptic form

of the enzyme, or in the transcript carrying alternative exon 5 and encoding the haematopoietic form of AChE (Fig. 5a). Thus, acute neuronal excitation mediated not only enhanced transcription, but also modified alternative splicing from the AChE gene, leading to *de novo* synthesis of the unique read-through AChE isoform.

In situ hybridization revealed low levels of read-through AChE mRNA in neuronal somata of cortical layers 2 and 5 of control mice. In contrast, both somata and apical dendrites of neurons from all cortical layers were intensely labelled for AChE mRNA following exposure to pyridostigmine (Fig. 5b). However, an exon-6-specific probe revealed similar levels of exon 6 AChE mRNA in somata of neurons in layers 2 and 5 of the parietal cortex of both control and inhibitor-treated mice (Fig. 5b, and data not shown). As otherwise non-AChE-expressing cells began producing large amounts of a secretable, non-synaptic form of AChE after acute cholinergic stimulation, non-cholinergic, non-catalytic activities could be attributed to read-through AChE in modulating long-term neuronal reorganization following cholinergic insults^{16–18}. This idea is consistent with reported homologies between AChE and a growing family of neuronal proteins involved in cell–cell interactions that include cytoplasmic domains capable of transducing signals¹⁹.

Our results indicate that modulated cholinergic gene expression acts to reduce available acetylcholine and depress cholinergic neurotransmission following stress. This molecular compensation would play a crucial role in short-term quietening of brain activity following a traumatic experience, but could have potentially damaging long-term implications. We previously reported that prolonged overexpression of AChE in the central nervous system of transgenic mice promotes delayed cognitive and neuroanatomical pathologies^{20–22}. Our present results therefore indicate a common mechanism for the delayed neuropsychiatric pathologies

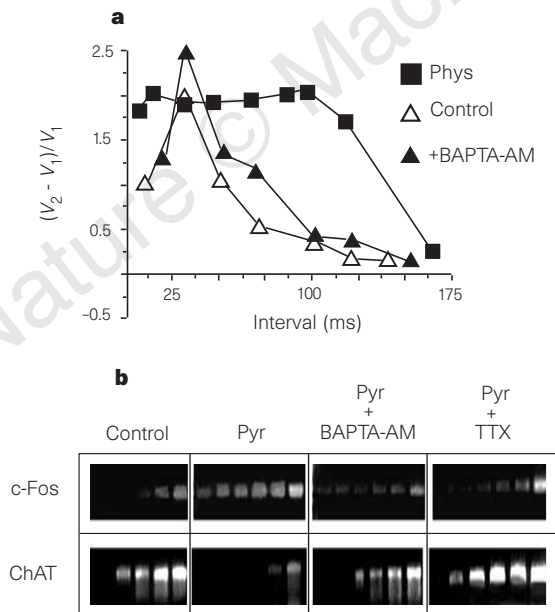


Figure 3 Physiological and transcriptional responses both depend on intracellular Ca^{2+} accumulation and Na^+ influx. **a**, Calcium chelator prevents enhancements. Paired-pulse facilitation was measured as in Fig. 2 in hippocampal slices under control conditions (open triangles), 1 h after the addition of $1 \mu\text{M}$ physostigmine (phys, filled squares) or 1 h after treatment with $1 \mu\text{M}$ physostigmine in the presence of the intracellular Ca^{2+} chelator BAPTA-AM ($1 \mu\text{M}$, filled triangles). BAPTA-AM alone had no effect on paired-pulse facilitation (data not shown). **b**, Suppression of the transcriptional response. c-Fos and ChAT mRNAs from control slices, or from slices treated for 1 h with 1 mM pyridostigmine (Pyr) alone or 1 mM Pyr with either BAPTA-AM ($1 \mu\text{M}$) or the Na^+ -channel-blocker tetrodotoxin (TTX, $1 \mu\text{M}$) were PCR-amplified as in Fig. 1a.

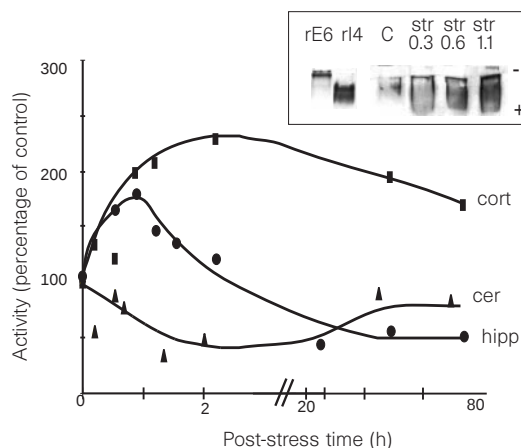


Figure 4 Long-term changes in AChE activity following stress. Average AChE activities (per cent of control) are shown in extracts of cortex (cort), cerebellum (cer), or hippocampus (hipp) from a total of 26 animals killed at the noted times after forced swimming. Specific activity (100%) in untreated controls was 3.98 ± 1.20 , 1.95 ± 1.34 , and 2.51 ± 0.39 nmol substrate hydrolysed per min per mg tissue for cortex, hippocampus and cerebellum, respectively. Inset shows that stress intensifies cortical AChE activity and diversifies its electrophoretic heterogeneity. Total proteins extracted from cortex of mice killed at the noted post-stress (str) times were electrophoresed ($20 \mu\text{l}$ per lane, $1:10$ wt/vol) in 7% non-denaturing polyacrylamide gels; catalytically active AChE was stained histochemically¹⁴. Recombinant human exon 6/AChE (rE6) and read-through AChE (read-through intron 4, rI4) produced in *Xenopus* oocytes ($1/5$ oocytes per lane) served as known correlates. C, control.

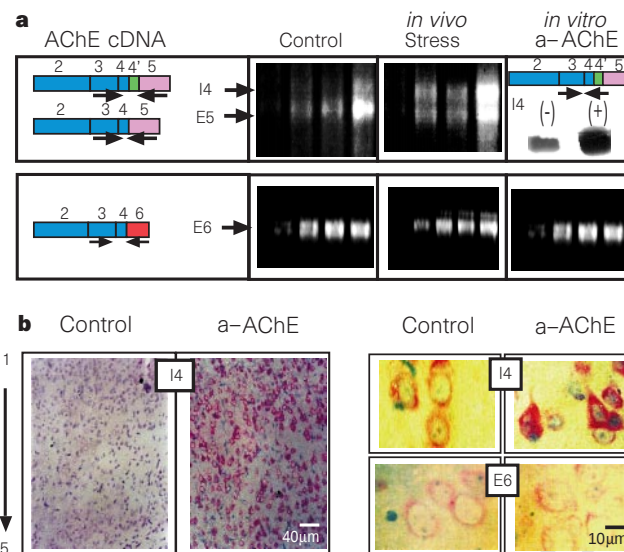


Figure 5 Selective induction of read-through AChE mRNA following stress and AChE inhibition. **a**, Kinetic follow-up of RT-PCR. Accumulated PCR products are derived from alternative AChE mRNAs under control, stress and anti-AChE (a-AChE) conditions. Positions of primer pairs specific for the read-through (intron 4, I4), synaptic (exon 6, E6) and erythrocyte (exon 5, E5) AChE mRNA subtypes are shown on the left. Endpoint products from a reaction using the I4-

specific primer pair +1,361 to -14 (74) were detected by hybridization with a radiolabelled nested probe. **b**, *In situ* hybridization. Cortical layers 1-5 (low magnification, left) and representative pyramidal neurons from cortical layer 2 (high magnification, right) are shown before and after *in vivo* exposure to pyridostigmine (a-AChE, 2 mg kg⁻¹). Stress treatment induced similar changes (not shown).

associated with PTSD and those associated with anticholinesterases. Feedback pathways leading to raised AChE may also be associated with low-level exposure to common anticholinesterase drugs and insecticides. Moreover, our discovery of anticholinesterase-promoted feedback mechanisms may help to explain the limited efficacy of cholinesterase inhibitors in treating Alzheimer's disease^{23,24} and the delayed neurocognitive disturbances reported by soldiers exposed to pyridostigmine or other anticholinesterases during the Gulf War^{25,26}.

Methods

Animal care and experimental protocols were carried out in accordance with institutional guidelines.

Forced swim stress protocol. This protocol was adapted for use in adult FVB/N mice as described⁷. Animals were subjected to two 4-min swim sessions in a water bath of 60 × 60 cm at 21 ± 1 °C.

Sagittal hippocampal brain slices. These slices were prepared and maintained as described²⁷, except that the concentration of KCl was 10 mM for molecular experiments and 5 mM for electrophysiological recordings. *Stratum oriens* fibres were stimulated with a bipolar tungsten electrode. Stimulation frequency was <0.3 Hz in all experiments. RT-PCR demonstrated the presence of stable mRNA for the key genes described for at least 12 h.

Kinetic follow-up of RT-PCR. This procedure was performed as described⁷ using the following selective primer pairs. Numbers indicate nucleotide positions in the Genebank cDNA sequences. For the AChE gene, nucleotides +375 to -1,160 were used to detect a region common to all AChE mRNA subtypes; nucleotides +1,361 to -1,896 (exon 6) were used to detect mRNA encoding synaptic AChE; and nucleotides +1,361 to exon 5 (-175) (intron 4/exon 5) were used to generate a 549-base-pair (bp) product from read-through AChE mRNA and a 432-bp product from mRNA encoding the erythrocyte-linked form of the enzyme. For ChAT mRNA, nucleotides +83 to -646 were used; for c-Fos mRNA, nucleotides +1,604 to -1,306 were used; and for synaptophysin mRNA, nucleotides +212 to -660 were used.

In situ hybridization. *In situ* hybridization was performed in 5 μm-thick paraffin-embedded cortical slices with 50-mer 5'-biotinylated, 2-O-methyl-protected complementary RNA probes²¹ beginning at intron I4 position 74 or exon E6 position 1,392 of the mouse AChE gene. Probes were detected using an

alkaline phosphatase/streptavidin conjugate and Fast Red (Boehringer) as a substrate. Counterstaining of nuclei was with Giemsa.

Received 24 November 1997; accepted 9 March 1998.

- Sapolsky, R. M. Why stress is bad for your brain. *Science* **273**, 749-750 (1996).
- Gurvits, T. V. et al. Neurological status of Vietnam veterans with chronic posttraumatic stress disorder. *J. Neuropsych. Clin. Neurosci.* **5**, 183-188 (1993).
- Rosenstock, L., Keifer, M., Daniell, W. E., McConnell, R. & Claypoole, K. Chronic central nervous system effects of acute organophosphate pesticide intoxication. *Lancet* **338**, 223-227 (1991).
- Wickelgr, I. The big easy serves up a feast to visiting neuroscientists: rat model for Gulf War Syndrome? *Science* **278**, 1404 (1998).
- Imperato, A., Puglisi-Allegra, S., Casolini, P. & Angellucci, L. Changes in brain dopamine and acetylcholine release during and following stress are independent of the pituitary-adrenocortical axis. *Brain Res.* **538**, 111-117 (1991).
- Ennis, M. & Shipley, M. T. Tonic activation of locus coeruleus neurons by systemic or intracerebral microinjection of an irreversible acetylcholinesterase inhibitor: increased discharge rate and induction of c-fos. *Exp. Neurol.* **118**, 164-177 (1992).
- Friedman, A. et al. Pyridostigmine brain penetration under stress enhances neuronal excitability and induces early immediate transcriptional response. *Nature Med.* **2**, 1382-1385 (1996).
- Ben Aziz-Aloya, R. et al. Expression of a human acetylcholinesterase promoter-reporter construct in developing neuromuscular junctions of *Xenopus* embryos. *Proc. Natl Acad. Sci. USA* **90**, 2471-2475 (1993).
- Li, Y., Camp, S., Rachinsky, T. L., Bongiorno, C. & Taylor, P. Promoter elements and transcriptional control of the mouse acetylcholinesterase gene. *J. Biol. Chem.* **268**, 3563-3572 (1993).
- Bausero, P. et al. Identification and analysis of the human choline acetyltransferase gene promoter. *Neuroreport* **4**, 287-290 (1993).
- Cervini, R. et al. Specific vesicular acetylcholine transporter promoters lie within the first intron of the rat choline acetyltransferase gene. *J. Biol. Chem.* **270**, 2465-2467 (1995).
- Cole, A. E. & Nicoll, R. A. Acetylcholine mediates a slow synaptic potential in hippocampal pyramidal cells. *Science* **221**, 1299-1301 (1983).
- Ghosh, A., Ginty, D. D., Bading, H. & Greenberg, M. E. Calcium regulation of gene expression in neuronal cells. *J. Neurobiol.* **25**, 294-303 (1994).
- Seidman, S. et al. Synaptic and epidermal accumulations of human acetylcholinesterase is encoded by alternative 3'-terminal exons. *Mol. Cell. Biol.* **15**, 2993-3002 (1995).
- Karpel, R. et al. Expression of three alternative acetylcholinesterase messenger RNAs in human tumor cell lines of different tissue origins. *Exp. Cell Res.* **210**, 268-277 (1994).
- Layer, P. G. & Willbold, E. Novel functions of cholinesterases in development, physiology and disease. *Prog. Histochem. Cytochem.* **29**, 1-99 (1995).
- Koenigsberger, C., Chiappa, S. & Brimijoin, S. Neurite differentiation is modulated in neuroblastoma cells engineered for altered acetylcholinesterase expression. *J. Neurochem.* **69**, 1389-1397 (1997).
- Sternfeld, M. et al. Acetylcholinesterase enhances neurite growth and synapse development through alternate contributions of its hydrolytic capacity, core protein and variable C-termini. *J. Neurosci.* **18**, 1240-1249 (1998).
- Ichchenko, K., Nguyen, T. & Sudhof, T. C. Structures, alternative splicing, and neuroligin binding of multiple neuroligins. *J. Biol. Chem.* **271**, 2676-2682 (1996).
- Beeri, R. et al. Transgenic expression of human acetylcholinesterase induces progressive cognitive deterioration in mice. *Curr. Biol.* **5**, 1063-1071 (1995).
- Andres, C. et al. AChE transgenic mice display embryonic modulations in spinal cord CHAT and neuroligin 1β gene expression followed by late-onset neuromotor deterioration. *Proc. Natl Acad. Sci. USA* **94**, 8173-8178 (1997).

22. Beeri, R. *et al.* Enhanced hemicholinium binding and attenuated dendrite branching in cognitively impaired AChE-transgenic mice. *J. Neurochem.* **69**, 2441–2451 (1997).
23. Knapp, M. J. *et al.* A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. *J. Am. Med. Assoc.* **271**, 985–991 (1994).
24. Winkler, M. A. Tacrine for Alzheimer's disease: which patient, what dose? *J. Am. Med. Assoc.* **271**, 1023–1024 (1994).
25. Sharabi, Y. *et al.* Survey of symptoms following intake of pyridostigmine during the Persian gulf war. *Isr. J. Med. Sci.* **27**, 656–658 (1991).
26. Milner, I. B. & Axelrod, B. N., Pasquantonio, J. & Sillanpaa, M. Is there a Gulf War syndrome? *J. Am. Med. Assoc.* **271**, 661 (1994).
27. Friedman, A. & Gutnick, M. J. Intracellular calcium and control of burst generation in neurons of guinea-pig neocortex in vitro. *Eur. J. Neurosci.* **1**, 374–381 (1989).

Acknowledgements. We thank M. J. Gutnick for inspiration and discussion, and M. Sternfeld for recombinant cholinesterase variants. This research was supported by grants from the US Army Medical Research and Development Command, the Israel Science Fund, the USA–Israel Binational Science Foundation, the Israeli Health Ministry and Ester/Medica Neuroscience (to H.S.). A.F. was a post-doctoral fellow of the Smith Psychobiology Fund and an incumbent of the Teva research prize for young investigators.

Correspondence and requests for materials should be addressed to H.S. (e-mail: soreq@shum.huji.ac.il).

Increased NMDA current and spine density in mice lacking the NMDA receptor subunit NR3A

Saumya Das*, Yasnory F. Sasaki*†, Thomas Rothe*†, Louis S. Premkumar§, Mari Takasu*, James E. Crandall||, Pieter Dikkes#, David A. Conner☆, Posina V. Rayudu‡, Wing Cheung‡, H.-S. Vincent Chen‡, Stuart A. Lipton‡† & Nobuki Nakanishi*†

* Department of Neurobiology, †Department of Genetics and Howard Hughes Medical Institute, and §Program in Neuroscience, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Cerebrovascular and Neuroscience Research Institute, Neurosurgical Service, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

§ Department of Physiology and Biophysics, State University of New York, Buffalo, New York 14214, USA

|| Developmental Neuroscience Division, The Shriver Center for Mental Retardation, Waltham, Massachusetts 02254, USA

Division of Neuroscience, Children's Hospital, Boston, Massachusetts 02115, USA

† These authors contributed equally to this work.

The NMDA (*N*-methyl-D-aspartate) subclass of glutamate receptor¹ is essential for the synaptic plasticity thought to underlie learning and memory^{2–4} and for synaptic refinement during development^{5,6}. It is currently believed that the NMDA receptor (NMDAR) is a heteromultimeric channel comprising the ubiquitous NR1 subunit and at least one regionally localized NR2 subunit^{7–11}. Here we report the characterization of a regulatory NMDAR subunit, NR3A (formerly termed NMDAR-L or χ -1), which is expressed primarily during brain development^{12,13}. NR3A co-immunoprecipitates with receptor subunits NR1 and NR2 in cerebrocortical extracts. In single-channel recordings from *Xenopus* oocytes, addition of NR3A to NR1 and NR2 leads to the appearance of a smaller unitary conductance. Genetic knockout of NR3A in mice results in enhanced NMDA responses and increased dendritic spines in early postnatal cerebrocortical neurons. These data suggest that NR3A is involved in the development of synaptic elements by modulating NMDAR activity.

We attempted to determine whether NR3A is a subunit of the NMDAR (Fig. 1a). First, using an anti-NR3A antibody, we defined the subcellular localization of NR3A in mouse cerebrocortical extracts at postnatal day 6 (P6). Like NR1, NR3A was selectively fractionated in a preparation enriched with postsynaptic densities (Fig. 1b). We next tested whether other glutamate-receptor subunits

could be co-immunoprecipitated with NR3A from cerebrocortical extracts. NR1 and NR2B were detected in anti-NR3A immunoprecipitates, but non-NMDA glutamate receptor subunits GluR2, GluR6/7 and delta1/2 were absent (Fig. 1c, lane 3). These results suggest that NR3A is associated with NMDAR subunits in brain extracts, although they do not constitute direct evidence that NR3A is part of the NMDAR channel.

We performed single-channel recordings from outside-out patches from *Xenopus* oocytes injected with NR1/NR2A ($n = 9$) or NR1/NR2A/NR3A ($n = 24$; Fig. 2). Our previous work¹⁴ had shown that expression of NR3A with NR1 and NR2 attenuated NMDA-evoked responses under two-electrode voltage-clamp^{12,13}. In our single-channel recordings, the presence of NR3A resulted in the appearance of a smaller unitary conductance (in 20 of 24 patches; Fig. 2b, d), in addition to the larger conductance previously observed for NR1/NR2A heteromers (Fig. 2a)¹⁴. Recordings were performed in low extracellular Ca^{2+} to minimize subconductance states. At -80 mV the larger state had a conductance of 75 ± 2.6 pS and the smaller 35 ± 3.4 pS (mean $\pm$ s.d. from amplitude histograms). Channels were analysed as described¹⁵. As the amount of injected NR3A cRNA was increased (ratio of

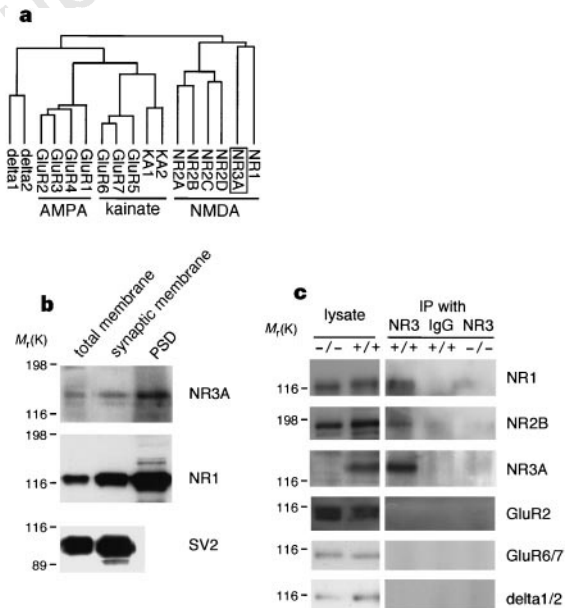


Figure 1 Biochemical characterization of NR3A. **a**, Alignment of members of the glutamate-receptor gene family based on amino-acid sequence homology. Progressive pairwise alignment of sequences²⁹ used the PILEUP program of the GCG software package for sequence analysis. Pharmacological classifications of subunits (see ref. 11 for functions of subunits other than NR3A) are shown under the alignment, and correlate well with groupings based on sequence similarities. The functional identification of NR3A as an NMDAR subunit extends this correlation. **b**, Subcellular localization of NR3A. Protein (25 μ g) was applied to each lane of 6% SDS-PAGE, and immunoblotting was performed using antibodies against NR3A, NR1 and the presynaptic protein SV2. The 'total membrane' fraction contained membrane proteins from synapses, endoplasmic reticulum and Golgi apparatus. PSD, postsynaptic density fraction. The validity of our fractionation was confirmed by the distribution of the control proteins, NR1 and SV2. **c**, Co-immunoprecipitation of NR1 and NR2B with NR3A from P6 brain extracts. Lane 1, NR3A^{-/-} extract; 2, wild-type extract; 3, immunoprecipitate (IP) with NR3A antibody from wild-type extract; 4, IP with rabbit anti-mouse IgG antibody from wild-type extract; and 5, IP with NR3A antibody from NR3A^{-/-} extract. Protein (25 μ g; lanes 1, 2) and IP from lysates initially containing 250 μ g protein (lanes 3–5) were subjected to immunoblotting with the antibodies shown on the right. Control experiments (lanes 4, 5) validated the specificity of both the immunoprecipitation procedure and the NR3A antibody.

Table 1 Measures of dendritic morphology for P19 cortical neurons

	Layer IV		Layer V	
	+/+ (n = 20)	-/- (n = 20)	+/+ (n = 15)	-/- (n = 15)
Basal dendrites				
Dendritic length (μm)	39.9 ± 3.0	47.6 ± 3.2	86.1 ± 5.6	88.2 ± 4.7
No. of primary branches	5.0 ± 0.2	5.3 ± 0.2	6.9 ± 0.3	6.8 ± 0.3
No. of branches	1.2 ± 0.1	1.7 ± 0.2	2.3 ± 0.2	2.3 ± 0.2
No. of spines	31.3 ± 2.3	104 ± 5.2	37.0 ± 2.7	110 ± 4.9
Spine density (spines per 100 μm)	78.4 ± 5.8	218 ± 11.0	43.0 ± 3.2	124 ± 5.3
Apical dendrites				
Dendritic length (μm)	55.4 ± 3.9	73.3 ± 5.1*	82.9 ± 6.8	95.3 ± 6.4
No. of primary branches	11.4 ± 1.1	12.2 ± 1.0	17.3 ± 1.1	18.9 ± 1.5
No. of spines	65.7 ± 6.4	222 ± 13.0	59.4 ± 5.1	225 ± 11.0
Spine density (spines per 100 μm)	118 ± 11.5	303 ± 18.3	71.7 ± 6.1	236 ± 12.0

Calculated or measured from camera lucida drawings of Golgi-stained neurons in layers IV and V of the neocortex of P19 male mice. Values are mean ± s.e.m. (3 animals for each genotype). Statistical significance was calculated by the *t*-test comparing values between wild-type (+/+) and NR3A-deficient (-/-) cells. Significance at the *P* < 0.05 level is indicated by an asterisk and at *P* < 0.001 by a solid box.

NR1:NR2A:NR3A increasing from 1:1:0 to 1:1:2), the frequency of occurrence of the large conductance state decreased from 0.26 ± 0.11 (mean ± s.d., *n* = 4) to 0.10 ± 0.08 (*n* = 4), while the mean open time stayed constant (3.5 ± 1.2 vs. 3.4 ± 0.4 ms). The decrease in absolute open probability of the large channel was accompanied by the appearance of a small channel with a frequency of occurrence of 0.05 ± 0.04 and a mean open time of 1.6 ± 0.3 ms.

No transitions were observed between the two conductance states in long-duration recordings (5–10 min), and in two patches only the small channel was seen (Fig. 2d). These observations strongly suggest that the two conductance states represent independent channels rather than subconductance levels. Similar results were obtained in patches recorded from oocytes injected with NR1/NR2B/NR3A cRNAs (data not shown). The smaller conductance

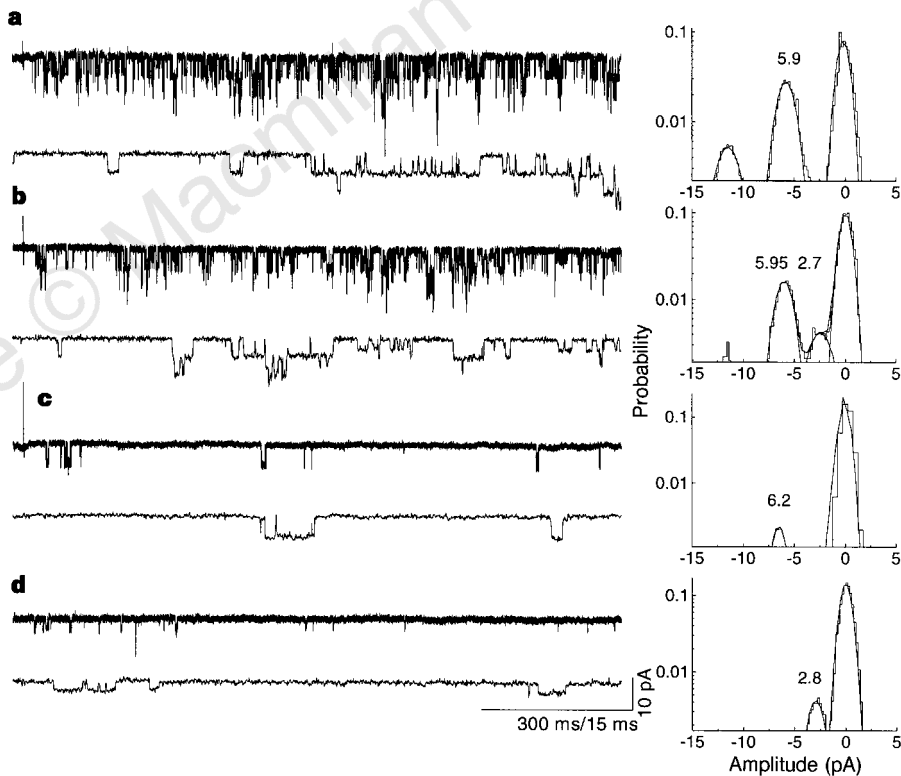


Figure 2 Single-channel recordings from outside-out patches expressing NR1, NR2A and NR3A. Currents were recorded from oocytes injected with cRNA encoding NR1/NR2A ± NR3A at a holding potential of -80 mV in response to application of 20 μM NMDA and 10 μM glycine (*n* = 33 patches); segments of traces at higher time resolution are illustrated below longer recordings. All points amplitude histograms (right) were constructed from a 5-s segment of representative data and fitted to gaussian components to show the size and frequency of occurrence of each channel. **a**, Recordings from an oocyte injected with NR1/NR2 (1:1). One predominant current level was observed with single and double openings (6.0 ± 0.2 pA or 75 ± 2.5; pS, mean ± s.d. for 9 patches), consistent with previous reports¹⁴. **b**, Recordings from an oocyte injected with NR1/NR2/NR3A

(1:1:2). As well as the main current level (~6 pA), a smaller current level was observed (2.8 ± 0.3 pA or 35 ± 3.8 pS for 4 patches). **c**, **d**, Representative recordings after injection of NR1/NR2A/NR3A (1:1:5) in 20 patches. In **c** (4 of 20 patches), only the large conductance channel was seen, but openings were infrequent. In **d** (2 of 20 patches), only the small conductance channels were activated in response to application of NMDA. No transitions to larger conductance channels were observed in recordings lasting at least 5–10 min. In other patches (14 of 20), injection of this higher proportion of NR3A cRNA reduced the number and the apparent open probability of the large conductance channels, and increased the relative frequency of small channel openings.

Table 2 Spine densities of P19 Purkinje cells and adult cerebrocortical neurons

	Dendritic spine density (spines per 100 μm)	
	+/+	-/-
P19 Purkinje proximal dendrites	80.8 $\pm$ 6.6 (<i>n</i> = 9)	79.3 $\pm$ 3.6 (15)
P19 Purkinje distal dendrites	77.7 $\pm$ 4.1 (6)	79.5 $\pm$ 4.3 (15)
Adult S1 layer IV apical dendrites	83.6 $\pm$ 5.0 (23)	101 $\pm$ 4.4 (18)

Dendritic spine densities (spines per 100 μm) of P19 Purkinje cells and adult S1 layer IV cerebrocortical neurons were calculated for wild-type (+/+) and NR3A-deficient (-/-) mice. Values are mean $\pm$ s.e.m. (*n* = 3 animals for each genotype). Significance at *P* < 0.05, calculated by *t*-test, is indicated by a box.

channels appeared to be less permeable to Ca^{2+} than the larger channels in the same patch, as manifested by a smaller shift in the reversal potential of the current-voltage relationship as the concentration of extracellular Ca^{2+} was raised to 10 mM (shift of ~ 14 mV for the larger channel against < 4 mV for the smaller channel). Taken together, these findings suggest that adding NR3A to NR1/NR2 subunits produces an NMDAR channel of smaller unitary conductance, shorter open time, and lower Ca^{2+} permeability.

To characterize function *in vivo*, we generated NR3A-deficient mice. The genomic region encoding the amino-terminal 118 amino-acid residues of NR3A was deleted by homologous recombination in embryonic stem (ES) cells (Fig. 3a). Homozygous offspring were born (Fig. 3b) with the expected Mendelian distribution. NR3A^{-/-} mice lacked both NR3A mRNA (Fig. 3c) and protein (Fig. 1c, NR3A immunoblot), and thus should be deficient in NR3A function. NR3A^{-/-} mice were fertile and grew to adulthood without apparent behavioural abnormalities. There were no discernible changes in the expression of NR1, NR2A and NR2B proteins (Fig. 3d). NR2C and NR2D mRNAs were not detected in the NR3A^{-/-} cerebral cortex at P7 by northern analysis (data not shown), a finding consistent with results from wild-type mice¹⁶.

We compared electrophysiological properties of acutely dissociated cerebrocortical neurons from P5–8 wild-type and NR3A^{-/-} mice. To determine the frequency of wild-type neurons expressing NR1 and NR3A mRNA in this preparation, we performed single-cell reverse transcriptase-based polymerase chain reactions (RT-PCR)¹⁷. NR3A PCR products were detected at high frequency among cells expressing NR1 (Fig. 3e). NMDA-evoked current measured by whole-cell recordings in NR3A^{-/-} neurons was qualitatively similar to that in wild-type cells: the current was blocked completely by D-AP5 and in a use-dependent manner by MK-801 (Fig. 3f). Cell capacitance, Na⁺ current density, and kainate-evoked current density were similar, whereas the NMDA-induced current density was 2.8-fold greater in NR3A^{-/-} neurons than in the wild type (Fig. 3g). The increased NMDA response in NR3A^{-/-} neurons is consistent with previous observations that NR3A attenuated NMDA responses of NR1/NR2 channels in oocytes^{12,13}. Further, single-channel recordings from outside-out patches of wild-type P8 cerebrocortical neurons revealed two conductance states similar to those in oocytes co-injected with NR1/NR2/NR3A cRNAs (data not shown). In contrast, the smaller conductance was not observed in outside-out patches from NR3A^{-/-} mice.

NMDAR activity has been implicated in the modification of dendritic arbors¹⁸ and spines^{19,20}. With enhanced NMDAR activity, neurons in NR3A^{-/-} mice might be expected to exhibit aberrant morphology. Haematoxylin and eosin staining, however, revealed no gross structural abnormalities in the cerebral cortex, barrel field, or other areas of the NR3A^{-/-} brain (data not shown). We next performed Golgi staining to delineate the dendritic morphology of cerebrocortical neurons at P19. NR3A mRNA expression has declined by P19 compared with younger states^{12,13}, but we still

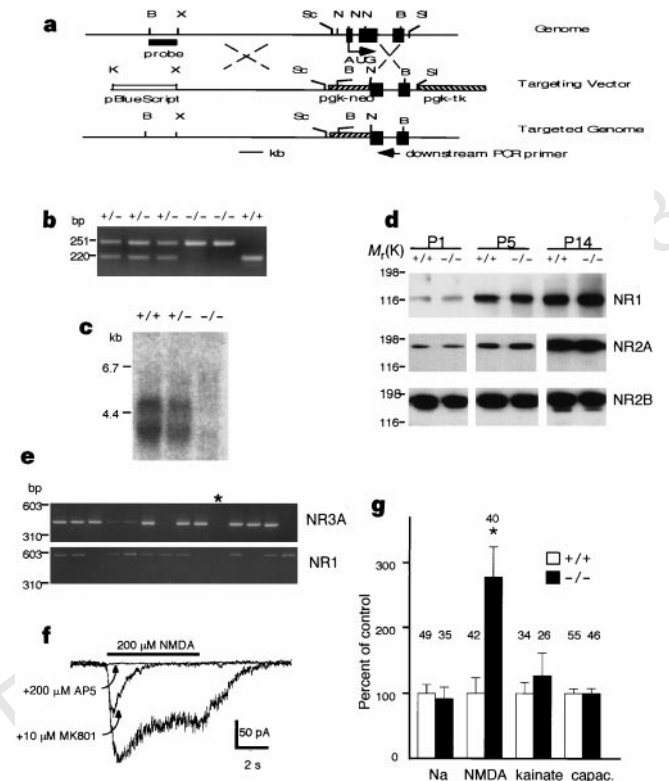


Figure 3 Biochemical and electrophysiological studies of NR3A^{-/-} mice.

a, Schematic representation of the endogenous NR3A locus, targeting vector and mutant allele following successful homologous recombination. Filled boxes represent NR3A exons. AUG denotes the translation initiation site. Dotted lines indicate sites of homologous recombination, which delete the entire first exon and a part of the second exon of the gene. Restriction enzymes used are: B, *Bgl*I; X, *Xho*I; Sc, *Sac*I; N, *Nar*I; Sl, *Sal*I; K, *Kpn*I. The probe for Southern hybridization and the downstream primer used for PCR analysis are indicated. **b**, PCR analysis of genotypes of mouse pups generated by intercrosses between two F₁ heterozygotes. PCR used genomic DNA from each pup in the presence of the common downstream primer and two upstream primers (one within the undisrupted NR3A genome DNA and the other in the pgk-neo sequence). The wild-type NR3A primer pair produced a 220-base-pair PCR product; the mutant generated a 251-bp DNA. **c**, Northern analysis of NR3A mRNA in mutant mice. Total brain mRNA (10 μg each) from P7 wild-type (+/+), heterozygotes (+/-) and homozygotes (-/-) were analysed using a radiolabelled DNA probe containing the entire NR3A coding region. A control experiment using the probe encoding glyceraldehyde-3-phosphate dehydrogenase showed that the amount of RNA loaded in each lane was similar (data not shown). **d**, Expression of NR1, NR2A and NR2B in NR3A^{-/-} mice. Immunoblotting was performed on cerebrocortical total membrane fractions containing 25 μg protein prepared from wild-type (+/+) and NR3A-deficient (-/-) mice at ages P1, P5 and P14. Filters were probed with antibodies against NR1, NR2A and NR2B. **e**, Single-cell RT-PCR analysis of acutely dissociated cerebrocortical neurons prepared from P7 wild-type mice. Each lane was loaded with PCR products derived from mRNA of a single cell. NR3A and NR1 PCR products were fractionated by 3% agarose gel electrophoresis. One of the 14 cells analysed (asterisk) yielded neither NR3A nor NR1 PCR products. **f**, Representative traces of NMDA-evoked currents in an acutely dissociated cerebrocortical neuron from a P8 NR3A^{-/-} mouse. Whole-cell currents recorded in the presence of 10 μM glycine in response to 200 μM NMDA, 200 μM NMDA with 10 μM MK-801, or 200 μM NMDA with 200 μM D-AP5. **g**, Current densities and whole-cell capacitance (capac.) of wild-type (+/+) and NR3A^{-/-} (-/-) cerebrocortical neurons at P5–8. Na, voltage-gated sodium current density (peak amplitude/capacitance); NMDA, NMDA (200 μM)-evoked current density; kainate, kainate (50 μM)-evoked current density. Data are shown as normalized means (wild-type values taken as 100%) $\pm$ s.e.m. The number of cells recorded is shown. Absolute values for wild-type neurons are: I_{Na} , 72 $\pm$ 9 pA pF⁻¹, I_{NMDA} , 1.21 $\pm$ 0.29 pA pF⁻¹, I_{kainate} , 5.04 $\pm$ 0.86 pA pF⁻¹, whole-cell capacitance, 12.6 $\pm$ 0.8 pF. **P* < 0.005 (Mann-Whitney *U* test; NR3A^{-/-} versus wild type).

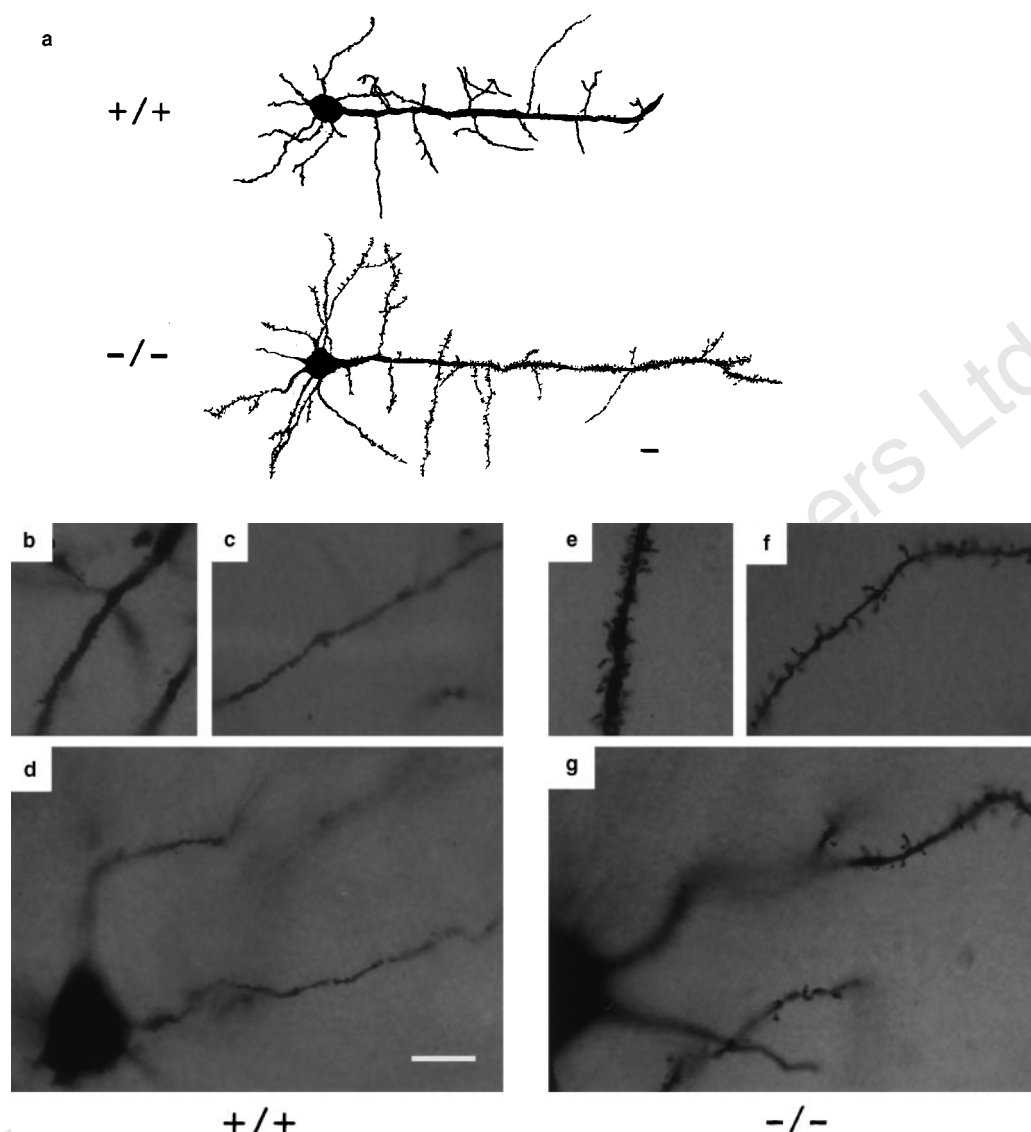


Figure 4 Changes in dendritic morphology of NR3A^{-/-} versus wild-type male mice at P19. **a**, Camera lucida drawings of Golgi-impregnated neurons in layer IV of the cerebral cortex from wild-type (+/+) and NR3A-deficient (-/-) mice. **b-g**, High-power views of the dendritic morphology of cerebrocortical layer IV neurons

revealed by Golgi staining in wild-type (**b-d**) and NR3A-deficient (-/-) (**e-g**) mice. Representative photographs (of 70 neurons) were taken to show morphologies of apical (**b, e**), lateral (**c, f**) and basal (**d, g**) dendrites. Scale bar, 10 μ m.

find detectable levels of NR3A protein in cerebrocortical brain extracts. From camera lucida drawings (Fig. 4a), we quantified the number of dendritic arbors and spines in cerebrocortical neurons of layers IV and V (Table 1; $n = 3$ mice for each genotype, a total of 70 neurons). The number and length of dendritic branches were largely unaltered in NR3A^{-/-} mice, although the length of apical dendrites was slightly elongated. The number of dendritic spines, however, was strikingly increased in NR3A^{-/-} mice. Spine morphology was also modified (Fig. 4b-g). Spine heads appeared enlarged and spine necks elongated in NR3A^{-/-} mice. As a control, neurons in an area where NR3A is not expressed (cerebellar Purkinje cells) were examined and no abnormality found (Table 2) in the same sagittal sections in which the dendritic morphology of cerebrocortical cells was affected. Thus morphological phenotypes observed in NR3A^{-/-} neurons are specific to cells normally expressing NR3A. Densities of dendritic spines were also quantified in adult mice for layer IV cerebrocortical neurons (Table 2; 3-month-

old mice, $n = 3$ mice for each genotype, a total of 41 neurons). The density of dendritic spines in NR3A^{-/-} mice was greater than in wild-type adults ($P < 0.014$; t -test). The difference, however, was less robust in the adult than in P19 mice (see Tables 1 and 2), possibly because expression of NR3A is greatly decreased in cerebrocortical neurons in adults compared to neonates^{12,13}.

Our findings suggest that NR3A functions as a regulatory subunit of NMDAR channels and is involved in the development of dendritic spines. Dendritic spines are sites where excitatory synapses are formed, and may be involved in cognitive function²¹ and neuroprotection²². Recent experiments suggest that dendritic-spine morphology is modified by excitatory activity^{19,20,23,24}. It is therefore possible that the increase in dendritic-spine density in NR3A^{-/-} mice may be caused by enhanced NMDAR-mediated synaptic input. Future studies of NR3A^{-/-} mice may help to determine the underlying biochemical pathway linking NMDAR activity to the modification of synaptic structures. □

Methods

Preparation of brain membrane fractions. Cortices were dissected and homogenized in 0.32 M sucrose, 10 mM Tris-HCl (pH 7.4). Nuclear fractions were removed from the homogenate by low-speed centrifugation. Recovered supernatant was subjected to high-speed centrifugation (16,000g for 1 h). The pellet was solubilized in a buffer (20 mM Tris-HCl (pH 7.4) 1 mM EDTA, protease inhibitors), and used as 'total membrane' fraction. Synaptic membrane and postsynaptic density-enriched fractions were prepared²⁵. The protein content of each fraction was quantified using the BioRad DC assay.

Antibodies. NR1 and GluR2 antibodies (Pharmingen) and rabbit anti-mouse IgG antibody (Jackson Immunochemical) were obtained. Anti-NR2A/NR2B antibodies and GluR2, delta1/2 and GluR6/7 antibodies were gifts from M. Sheng and R. Wenthold, respectively, and SV2 antibody was obtained from K. M. Buckley. NR3A antibody was prepared as follows. cDNA encoding a portion (amino acids 780–914) of the NR3A protein was inserted into the *Escherichia coli* expression vector pGex-4T-3 to produce a fusion protein between glutathione S-transferase (GST) and the NR3A sequence. The fusion protein was injected into rabbits to generate antisera, from which antibody was purified by affinity chromatography. The purified antibody detects a prominent band at relative molecular mass ~130,000 (M_r 130K) (Fig. 1b, c), which is consistent with the calculated M_r (124.5K) of NR3A.

Immunoprecipitation and immunoblotting. Brain extracts were treated with 2% SDS before immunoprecipitations, which were carried out by incubating lysates with primary antibody (10 μ g ml⁻¹) at 4 °C for 1 h. Immunoprecipitates were eluted from protein A sepharose in SDS sample buffer, and immunoblotting was performed using peroxidase-conjugated anti-rabbit IgG (Jackson Immunochemical) and enhanced chemiluminescent reagents (Amersham).

Generation of NR3A^{-/-} mice. C1 ES cells prepared from 129/SvJ mice were transfected with the targeting vector using electroporation. Homologous recombination was observed in 5 of 350 neo-resistant ES cells. ES cells with the disrupted NR3A locus were injected into C57BL/6-derived blastocysts to generate chimaeric mice. Male chimaeric mice were then crossed with BlackSwiss or 129SvEv female mice to produce F₁ heterozygotes. For most experiments, mice with 129Sv/BlackSwiss backgrounds were used, and selected electrophysiological experiments were also performed using mice with the 129Sv background. With regard to the electrophysiological properties of cerebrocortical neurons, no significant differences were observed between the two genetic backgrounds.

Genomic PCR. Primer sequences and PCR conditions will be provided upon request.

Preparation of cerebrocortical neurons. Tissues containing parieto-occipital cortex were incubated twice (20 min each) at 37 °C in Hanks' balanced salt solution (HBSS), pH 7.22, containing 3.5 U ml⁻¹ papain, 1.7 mM cysteine, 20 μ g ml⁻¹ bovine serum albumin and NMDAR blockers (100 μ M DL-AP5, 1 mM kynurenic acid, 10 mM MgCl₂). After rinsing, tissues were gently triturated with a 1-ml glass serological pipette. The supernatant was then placed onto glass coverslips coated with poly-L-lysine. Cells were plated 2–10 h before electrophysiological recording.

Single-cell RT-PCR. Cytoplasmic contents were collected from neurons and reverse transcribed¹⁷. PCR primers were taken from two separate exons of NR1 (ref. 26) and NR3A. Sequences for two pairs of nested primers for NR3A were: 5'-GGTCCACCCGGCTCCGAAAG-3' and 5'-TTGGAGTTATCTCCGTA-AGGAGGAGGAA-3' (outside primers) and 5'-CCGCGGGATGCCCTACT-GTTC-3' and 5'-CCAGTTGTTTCATGGTCAGGAT-3' (inside primers). Primer sequences for NR1 were 5'-ACACAGGAGCGGGTAAACAAC-3', and 5'-CTGTCTGAGGGGTTCTGAG-3'. Each PCR was carried out in 1× PCR buffer I (Perkin-Elmer) for 40 cycles, and each cycle consisted of 95 °C for 1 min, 62 °C for 1 min, and 72 °C for 1 min.

Electrophysiological recording. Whole-cell and single-channel currents from outside-out patches were recorded from cerebrocortical neurons at 20–25 °C by an experimenter blinded to genotype, as described^{27,28}. Electrodes were filled with an intracellular solution (pH 7.22) containing (in mM): 120 CsCl, 20 tetraethylammonium-Cl, 10 HEPES, 2.25 EGTA, 1 CaCl₂ and 2 MgCl₂. The external solution was a modified HBSS with 10 μ M glycine in which Mg²⁺ salts were omitted, and [Ca²⁺]_o was adjusted to 2.5 mM or lowered to approximately the nanomolar range by adding 1 mM EGTA in the absence of added CaCl₂.

Single-channel recordings from oocytes were performed on outside-out patches¹⁵. The intracellular solution contained (in mM): 90 Na-gluconate, 10 NaCl, 4 ATP, 0.25 GTP, 10 BAPTA, and the pH was adjusted to 7.3 with NaOH. The extracellular solution contained (in mM): 100 NaCl, 2.5 KCl, 5 HEPES, 1.5 EGTA, and the pH was adjusted to 7.4 with NaOH. EGTA was omitted when [Ca²⁺]_o was raised to 10 mM.

Histological procedures. Transverse sections were analysed after Golgi staining. Values in Tables 1 and 2 were obtained from camera lucida drawings performed under light microscopy. For the cerebral cortex, neurons in the S1 area of layers IV and V were traced. Apical dendrites were defined as single, thick dendrites that extended from the top of the neuronal soma towards layer I; basal dendrites were defined as extending radially from the soma. For cerebellar Purkinje cells, secondary and tertiary branches were traced for each of the neurons analysed, and spine density was calculated for proximal and distal levels separately. A *t*-test was performed to compare the values obtained for wild-type and NR3A^{-/-} neurons.

Received 30 December 1997; accepted 4 March 1998.

- McBain, C. J. & Mayer, M. L. N-methyl-D-aspartic acid receptor structure and function. *Physiol. Rev.* **74**, 723–760 (1994).
- Gustafsson, B. & Wigstrom, H. Physiological mechanisms underlying long-term potentiation. *Trends Neurosci.* **11**, 156–162 (1988).
- Madison, D. V., Malenka, R. C. & Nicoll, R. A. Mechanisms underlying long-term potentiation of synaptic transmission. *Annu. Rev. Neurosci.* **14**, 379–397 (1991).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Constantine-Paton, M., Cline, H. T. & Debski, E. Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Annu. Rev. Neurosci.* **13**, 129–154 (1990).
- Singer, W. Development and plasticity of cortical processing architectures. *Science* **270**, 758–764 (1995).
- Moriyoshi, K. *et al.* Molecular cloning and characterization of the rat NMDA receptor. *Nature* **354**, 31–37 (1991).
- Kutsuwada, T. *et al.* Molecular diversity of the NMDA receptor channel. *Nature* **358**, 36–41 (1992).
- Monyer, H. *et al.* Heteromeric NMDA receptors: molecular and functional distinction of subtypes. *Science* **256**, 1217–1221 (1992).
- Sheng, M., Cummings, J., Roldan, L. A., Jan, Y. N. & Jan, L. Y. Changing subunit composition of heteromeric NMDA receptors during development of rat cortex. *Nature* **368**, 144–147 (1994).
- Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
- Sucher, N. J. *et al.* Development and regional expression pattern of a novel NMDA receptor-like subunit (NMDAR-L) in the rodent brain. *J. Neurosci.* **15**, 6509–6520 (1995).
- Ciabarra, A. M. *et al.* Cloning and characterization of χ -1: a developmentally regulated member of a novel class of the ionotropic glutamate receptor family. *J. Neurosci.* **15**, 6498–6508 (1995).
- Stern, P., Miroslav, C., Colquhoun, D. & Stephenson, F. A. Single channel properties of cloned NMDA receptors in a human cell line: comparison with results from *Xenopus* oocytes. *J. Physiol. (Lond.)* **476**, 391–397 (1994).
- Premkumar, L. S. & Auerbach, A. Stoichiometry of recombinant N-methyl-D-aspartate receptor channels inferred from single-channel current patterns. *J. Gen. Physiol.* **110**, 485–502 (1997).
- Monyer, H., Burnashev, N., Laurie, D. J., Sakmann, B. & Seeburg, P. H. Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron* **12**, 529–540 (1994).
- Lambole, B., Audinat, E., Bochet, P., Crepel, F. & Rossier, J. AMPA receptor subunits expressed by single Purkinje cells. *Neuron* **9**, 247–258 (1992).
- McAllister, A. K., Katz, L. C. & Lo, D. C. Neurotrophin regulation of cortical dendritic growth requires activity. *Neuron* **17**, 1057–1064 (1996).
- Woolley, C. S. & McEwen, B. S. Estradiol regulates hippocampal dendritic spine density via an N-methyl-D-aspartate receptor-dependent mechanism. *J. Neurosci.* **14**, 7680–7687 (1994).
- Woolley, C. S., Weiland, N. G., McEwen, B. S. & Schwartzkroin, P. A. Estradiol increases the sensitivity of hippocampal CA1 pyramidal cells to NMDA receptor-mediated synaptic input: correlation with dendritic spine density. *J. Neurosci.* **17**, 1848–1859 (1997).
- Harris, K. M. & Kater, S. B. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annu. Rev. Neurosci.* **17**, 341–371 (1994).
- Segal, M. Dendritic spines for neuroprotection: a hypothesis. *Trends Neurosci.* **18**, 468–471 (1995).
- Papa, M. & Segal, M. Morphological plasticity in dendritic spines of cultured hippocampal neurons. *Neuroscience* **71**, 1005–1011 (1996).
- Kossel, A., Williams, C. V., Scheizer, M. & Kater, S. B. Afferent innervation influences the development of dendritic branches and spines via both activity-dependent and non-activity-dependent mechanisms. *J. Neurosci.* **17**, 6314–6324 (1997).
- Rogers, S. W. *et al.* The characterization and localization of the glutamate receptor subunit GluR1 in the rat brain. *J. Neurosci.* **11**, 2713–2724 (1991).
- Hollmann, M. *et al.* Zinc potentiates agonist-induced currents at certain splice variants of the NMDA receptor. *Neuron* **10**, 943–954 (1993).
- Chen, H.-S. V. *et al.* Open-channel block of NMDA responses by memantine: therapeutic advantage against NMDA receptor-mediated neurotoxicity. *J. Neurosci.* **12**, 4427–4436 (1992).
- Chen, H.-S. V. & Lipton, S. A. Mechanism of memantine block of NMDA-activated channels in rat retinal ganglion cells: uncompetitive antagonism. *J. Physiol. (Lond.)* **499**, 27–46 (1997).
- Sucher, N. J., Awobuluyi, M., Choi, Y.-B. & Lipton, S. A. NMDA receptors: from genes to channels. *Trends Pharmacol. Sci.* **17**, 348–355 (1996).

Acknowledgements. We thank N. J. Sucher, M. Wyszynski, D. Zurakowski, J. Yuan, T. S. Khurana, C. Yiannoutsos and K. Rosen for advice and discussion; N. J. Sucher, M. Sheng, R. Wenthold and K. Buckley for antibodies; and J. Cohen and M. Sheng for critical reading of the manuscript. This work was supported by fellowships from the HHMI (S.D., W.C.) and the Harvard Mahoney Neuroscience Institute (Y.F.S.), grants from the NIH (J.E.C., S.A.L., N.N.), the Klingenstein Foundation, the Edward R. and Anne G. Lefler Center, NARSAD and the Funds for Discovery (N.N.). S.A.L. has consultancy and sponsored research agreements with Neurobiological Technologies, Inc. (Richmond, CA) and Allergan, Inc. (Irvine, CA) in the field of NMDAR antagonists.

Requests for the NR3A cDNA clone or antibody should be addressed to S.A.L. (e-mail: slipton@rics.bwh.harvard.edu). All other correspondence should be addressed to N.N. (e-mail: nnakanis@warren.med.harvard.edu).

Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain

Eric H. Schroeter, Jeffrey A. Kisslinger & Raphael Kopan

Division of Dermatology and the Department of Molecular Biology and Pharmacology, Washington University, Box 8123, 4940 Parkview Place, St Louis, Missouri 63110, USA

Notch proteins are ligand-activated transmembrane receptors involved in cell-fate selection throughout development¹⁻³. No known enzymatic activity is contained within Notch and the molecular mechanism by which it transduces signals across the cell membrane is poorly understood. In many instances, Notch activation results in transcriptional changes in the nucleus through an association with members of the CSL family of DNA-binding proteins (where CSL stands for CBF1, Su(H), Lag-1)¹⁻⁴. As Notch is located in the plasma membrane and CSL is a nuclear protein, two models have been proposed to explain how they interact (Fig. 1). The first suggests that the two interact transiently at the membrane^{1,5-7}. The second postulates that Notch is cleaved by a protease, enabling the cleaved fragment to enter the

nucleus^{6,8-14}. Here we show that signalling by a constitutively active membrane-bound Notch-1 protein requires the proteolytic release of the Notch intracellular domain (NICD), which interacts preferentially with CSL. Very small amounts of NICD are active, explaining why it is hard to detect in the nucleus *in vivo*. We also show that it is ligand binding that induces release of NICD.

Earlier work showed that Notch is proteolytically cleaved¹³. To determine whether this cleavage is essential for Notch-1 signalling, we identified and mutated the cleavage site to alter Notch-1 processing. To generate enough NICD for sequencing, we over-expressed a version of constitutively activated membrane-bound Notch-1 (N^{ΔE}) (see Fig. 2a and Methods) in 10T1/2 cells using a Sindbis virus expression vector¹⁵. The two bands of lower relative molecular mass (Fig. 2b) were purified and subjected to ten cycles of amino-terminal sequencing. The amino-acid sequence VLLSRKRRRQHG was obtained from both bands, indicating that NICD is generated as a result of proteolytic cleavage between amino acids G1743 and V1744 (Fig. 2a). The difference in the size of the two bands is explained by the observation that NICD seems to be modified after cleavage (Fig. 2b), probably by phosphorylation (E.H.S. and R.K., unpublished observations). The sequence surrounding

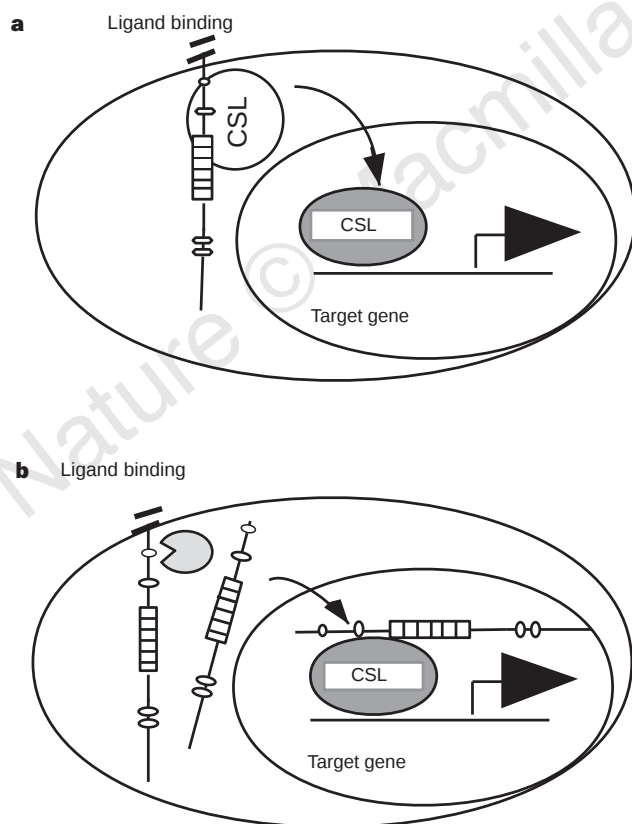


Figure 1 Different predictions are made by two models of Notch signalling. **a**, Models not invoking processing propose that interactions between Notch and CSL at the membrane may be sufficient to transduce a signal. Thus, ligand-regulated NICD release is not expected to be necessary for signalling. **b**, The processing model suggests that Notch signalling requires the release of NICD, which is capable of direct interaction with CSL in the nucleus, to turn on transcription of target genes. This model predicts that NICD release is regulated by ligand binding and that blocking proteolysis interferes with signalling. Notch-1 domain symbols as in Fig. 2.

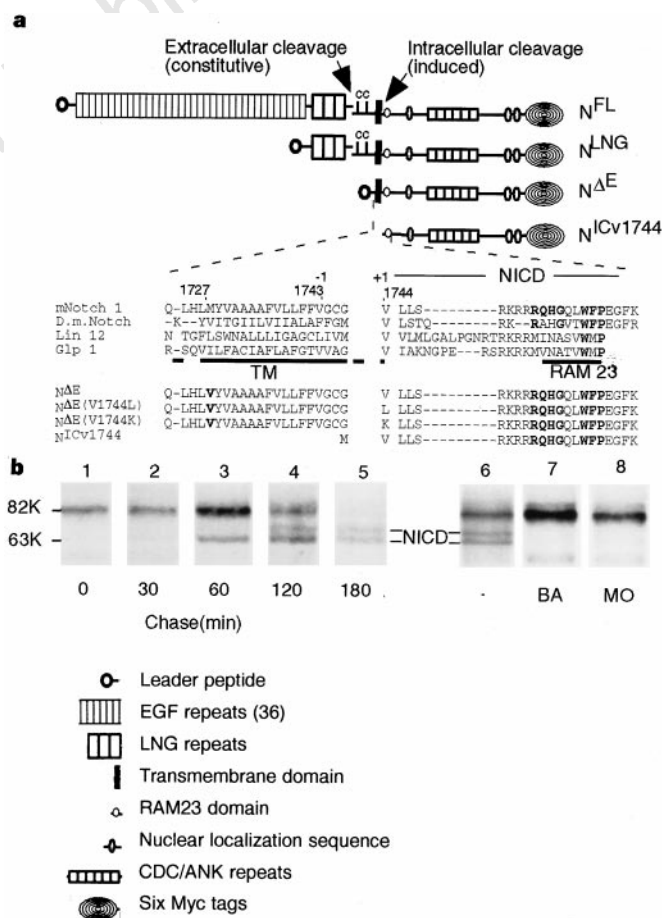


Figure 2 The Notch-1 constructs used and processing of N^{ΔE} to produce NICD. **a**, Map of domains and cleavage sites found in Notch-1 constructs used in this study. Cleavage occurs at or near the transmembrane domain (TM) between G1743(-1) and V1744(+1) (PIR entry A46019 (www-nbrf.georgetown.edu/pir/)). EGF, epidermal growth factor. M, murine; D.M., *Drosophila melanogaster*. **b**, Characterization of N^{ΔE} cleavage and release of NICD. Both lower bands in lanes 4-6 have the same N-terminal end and both represent NICD. Pulse-chase analysis shows that the lower band is produced first; a form with higher relative molecular mass appears later (lanes 3-5). Processing is inhibited by brefeldin A (BA; lane 7) and monensin (MO; lane 8), indicating that cleavage occurs after passage through the *trans*-Golgi network. Lane 6, untreated control.

G1743 and V1744 has no identity with known protease sites. Furthermore, V1744 seems to be conserved in all known Notch proteins and resides within or very close to the cytoplasmic side of the transmembrane domain (Fig. 2a).

Cleavage of Notch-1 to produce NICD is inhibited by brefeldin A (Fig. 2b, lane 7), which disassembles the Golgi apparatus^{16,17}, and by monensin (Fig. 2b, lane 8), which prevents transfer of proteins from the Golgi to the plasma membrane^{18,19}. Thus, proteolytic release of NICD apparently occurs after export of Notch-1 from the Golgi and may take place at the plasma membrane or after internalization of membrane-bound Notch-1 protein. Kuzbanian (Kuz), an ADAM family protease^{20–22}, has been proposed to mediate cleavage of Notch-1 at or near a previously identified site in the extracellular domain¹³ (Fig. 2a). However, this constitutive cleavage is monensin-resistant and may take place as part of the maturation of Notch in the secretory pathway²³. In addition, dominant-negative Kuz²⁰ does not affect NICD release or N^{ΔE} activity in our system (data not shown), indicating that the cleavage releasing NICD is unlikely to be mediated by Kuz.

Notch signalling can initiate expression of the repressor HES-1. We tested the importance of proteolysis for Notch1-mediated regulation of the HES-1 promoter by mutating V1744 (Fig. 2a). If proteolysis is necessary for Notch-1 activity, mutations that alter cleavage of Notch-1 should correspondingly alter its activity. We evaluated effects on transcription by measuring steady-state levels of luciferase expressed from the HES-1 promoter over a range of concentrations of Notch-1 plasmid DNA in transiently transfected 3T3 cells. N^{ΔE(V1744)} mutants show a reduction in the amount of NICD produced and a parallel reduction in activity (Fig. 3a, lanes

6–8; Fig. 4a, mutants N^{ΔE(V1744L)} and N^{ΔE(V1744K)}; see Methods). Differential stability is not responsible for this effect as equivalent levels of unprocessed Notch-1 protein are expressed from all mutant constructs over several different plasmid concentrations (Fig. 3a, and data not shown). As positive controls, N^{ΔE} and a construct in which V1744 becomes the first amino acid following the initiating methionine (N^{ICV1744}) were transfected. An inactive Notch-1 protein containing LNG repeats (Fig. 2a, N^{LNG})¹³ was a negative control. Although N^{ΔE(V1744)} mutant proteins that are poorly processed have significant activity at 10 ng of plasmid, N^{ΔE} produces the same level of activity at much lower concentrations (Fig. 4a).

These results implicate NICD as the activator of HES-1 expression. However, it is possible that the effect of V1744 mutations on Notch-1 activity may be independent of cleavage. If this is true, nuclear Notch-1 proteins containing the V1744 mutations should also have reduced activity. V1744 mutant proteins that lack the transmembrane domain (N^{IC(V1744L)} and N^{IC(V1744K)}) perform as well as N^{ΔE} and N^{ICV1744} in activation of the HES-1 promoter (Fig. 4b). These results establish that Notch-1 proteins with mutations of V1744 only exhibit reduced activity when membrane-bound (Fig. 4).

The attenuation of N^{ΔE} activity caused by cleavage mutations indicates the importance of proteolysis in Notch signalling. However, V1744 is located only seven amino acids away from the conserved RAM23 site (Fig. 2a) required for *in vitro* binding of Notch to CSL^{6,12,14}. This proximity raises the concern that these mutations may interfere with interactions between Notch and CSL at the membrane. We therefore examined the ability of the CSL protein RBP3¹¹ (CSL^{RBP3}) to bind forms of N^{ΔE} in which the cleavage site was altered. In tissue culture cells, CSL^{RBP3} precipitates equally

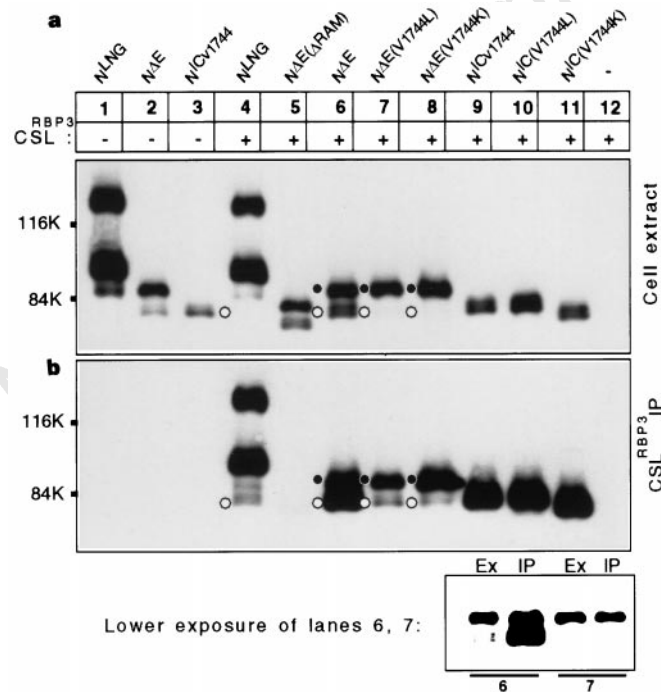


Figure 3 Western blot analysis. **a**, Equivalent amounts of unprocessed N^{ΔE} and N^{ΔE(V1744)} mutant proteins (filled circles) are detected (lanes 2, 6–8). In N^{ΔE(V1744)} mutants, production of NICD (open circles, lanes 7, 8) is greatly reduced. **b**, Notch-1 precipitated together with SL^{RBP3} from extracts used in (a). N^{ΔE} and N^{ΔE(V1744)} mutants all precipitate with equal efficiency (lanes 6–8, filled circles); however, NICD is enriched (open circles and inset). N^{ICV1744} and its mutants all precipitate with equal efficiency (lanes 9, 10). Notch-1 is not precipitated in the absence of Flag-CSL^{RBP3} (lanes 1–3) or when the RAM23 domain is deleted (lane 5). No cross-reactivity to Myc is found in the absence of tagged Notch-1 (lane 12). IP, immunoprecipitate; Ex, extract.

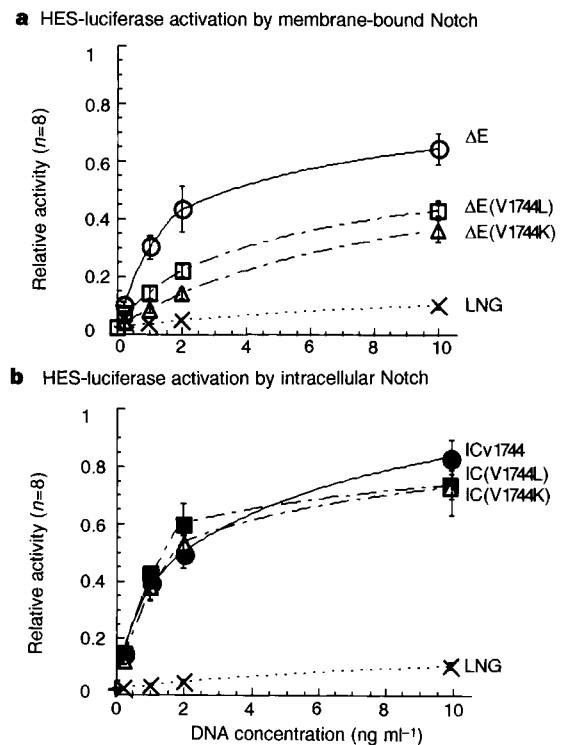


Figure 4 Activation of HES-1-luc by Notch-1 proteins. **a**, The mutant N^{ΔE(V1744)} molecules activate the HES-1 promoter at a significantly slower rate and reach <60% of the maximum activity of N^{ΔE}. This level of activity (60% of maximum) is reached by N^{ΔE} at tenfold lower DNA concentration. **b**, N^{ICV1744} and its mutants all accumulate luciferase at an equal rate and have the same maximum activity. Data are normalized as percentage of the activation seen with 100 ng ml⁻¹ of N^{ICV1744} plasmid DNA, and the results are averaged for each data point (n = 8). Standard errors are shown.

with the membrane-bound fraction of all $N^{\Delta E}$ proteins (Fig. 3a, b, lanes 6–8, dark dots). The same is found for all the $N^{ICV1744}$ constructs (Fig. 3a, b; lanes 9–11). Significantly, CSL^{RBP3} preferentially interacts with NICD in these cell extracts. Although only a small fraction of $N^{\Delta E}$ is processed, most of the protein precipitated with CSL^{RBP3} corresponds to NICD (Fig. 3, inset). Similar enrichment of NICD is seen with the $N^{\Delta E(V1744)}$ mutants. The amounts of NICD produced in cells expressing these proteins correlates with their activity, as follows: $N^{ICV1744} > N^{\Delta E} > N^{\Delta E(V1744K)} \geq N^{\Delta E(V1744L)} > N^{LNG}$.

Although CSL proteins do not appear to localize with Notch together in the cytoplasm *in vivo*^{24,25}, transient interactions may be occurring at the membrane. However, CSL preferentially binds NICD in transfected cells (Fig. 3a, b, lanes 6–8), whereas *in vitro*-translated $N^{ICV1744}$ and $N^{\Delta E}$ interact with CSL^{RBP3} with comparable affinity (not shown). Moreover, DNA-binding N^{IC}/CSL^{RBP3} complexes are found in nuclear extracts¹¹. These results indicate that processing of Notch-1 at V1744 is required to release an intracellular fragment to the nucleus, where the fragment acts with CSL to activate transcription. Inhibiting this step with mutations affects their activity.

The $N^{\Delta E(V1744)}$ mutants and N^{LNG} do exhibit some activity. Therefore NICD must act at very low nuclear concentrations. To study this possibility, we first estimated how much protein is produced from pCS2⁺ expression vectors by titrating pCS2⁺ β -galactosidase DNA over a range of concentrations. β -Galactosidase protein accumulates linearly as a function of pCS2⁺ DNA concentration (Fig. 5a). From 0.2 to 2 ng ml⁻¹ of DNA, we detect 0.2–2 pg β -galactosidase protein per extract. These values are just above detectable levels in this assay. The HES-1 promoter, however, responds logarithmically, saturating at concentrations above

10 ng ml⁻¹ pCS2⁺ $N^{ICV1744}$ DNA (Figs 4b and 5a). This implies that by the time that detectable amounts of $N^{ICV1744}$ have accumulated, the response to $N^{ICV1744}$ has already occurred and is at >50% of saturation.

Next we estimated amounts of NICD required in single cells to produce a response from a HES-1- β -galactosidase reporter. Activation of this reporter occurred at concentrations below our Notch-1 detection level. Accumulation of $N^{ICV1744}$ protein was monitored by antibody staining in transiently transfected cells over a range of pCS2⁺ $N^{ICV1744}$ DNA concentrations. At the same time, HES-1 promoter activity was scored by staining for β -galactosidase (Fig. 5). To increase the sensitivity of Notch-1 detection, the OPA/PEST domains were replaced with six Myc tags¹³. Using the anti-Myc antibody 9E10, $N^{ICV1744}$ can be detected at DNA concentrations starting at 2 ng ml⁻¹, indicating that detection by this antibody is similar in sensitivity to the enzymatic assay used for β -galactosidase. The number of cells expressing HES-1- β -galactosidase increases as the pCS2⁺ $N^{ICV1744}$ DNA concentration increases (Fig. 5b). However, $N^{ICV1744}$ at a concentration of 2 ng ml⁻¹ is detected in only 2% of the immunoreactive cells and no $N^{ICV1744}$ is found in the cells receiving less pCS2⁺ $N^{ICV1744}$ DNA (Fig. 5b, d). It is only at higher Notch-1 DNA concentrations (20–100 ng; Fig. 5b, e) that most HES-1- β -galactosidase positive cells express detectable amounts of $N^{ICV1744}$ protein. Moreover, even at these higher DNA concentrations, cells expressing only β -galactosidase are found in numbers significantly higher than background (Fig. 5b). These cells must also be expressing undetectable levels of Notch-1. These data confirm that the activation threshold of the HES-1 promoter is at very low nuclear concentrations of Notch-1 and may explain why nuclear NICD has not been detected by antibodies in developing tissues.

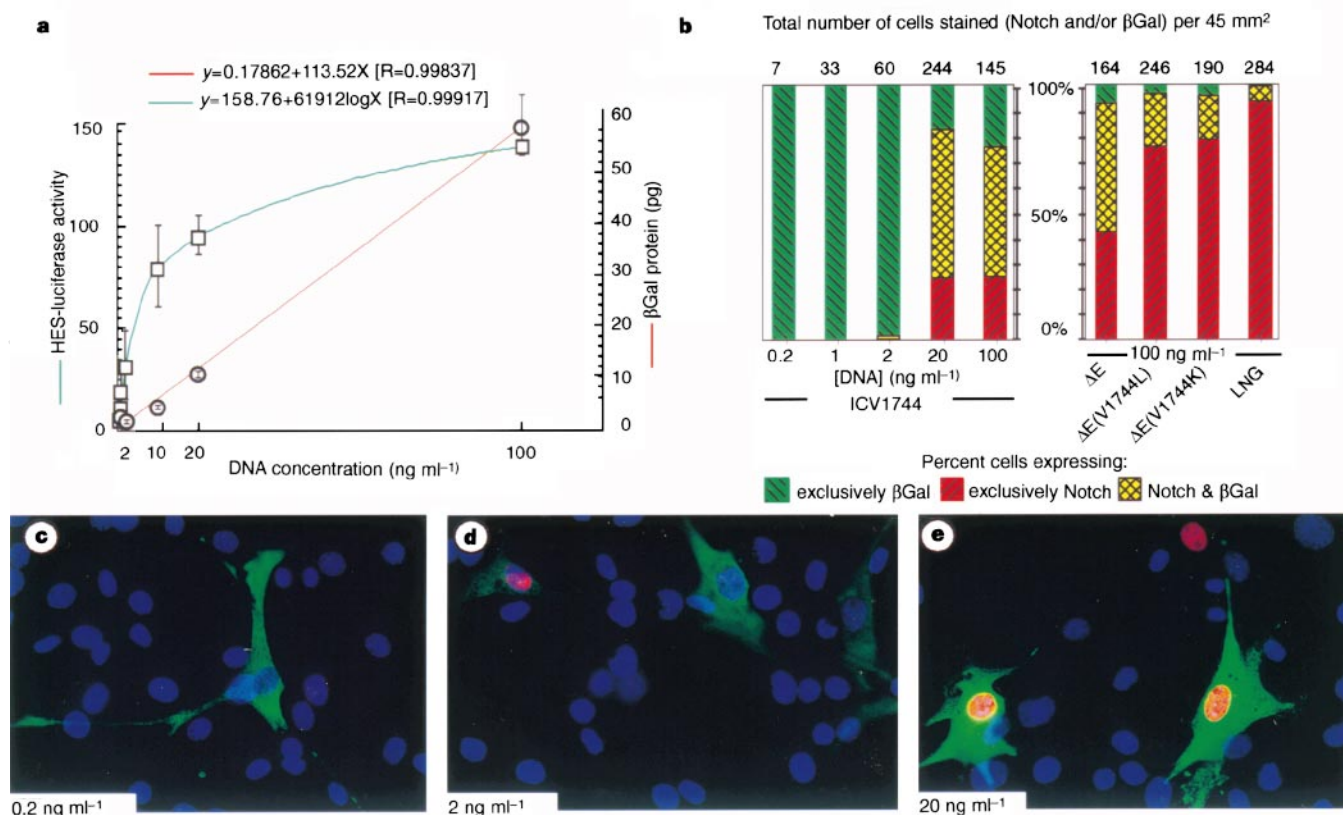


Figure 5 Notch-1 acts below threshold of detection. **a**, Production of β -galactosidase (β Gal) protein from the pCS2⁺ vector is linear. Production of $N^{ICV1744}$ from pCS2⁺ induces a logarithmic response from the HES-1 promoter. **b**, pCS2⁺ $N^{ICV1744}$ titrated in the presence of the HES-1- β Gal reporter, monitored by immunofluorescence. Total numbers of stained cells (~20,000 scanned) are

shown above each bar. The threshold of HES-1 activation is lower than the threshold of detection of Notch-1. The percentage of β Gal-positive cells decreases in cells expressing mutant $N^{\Delta E(V1744)}$ or N^{LNG} . **c–e**, High-magnification view of $N^{ICV1744}$ transfected cells described in (**b**). Red, Notch-1; green, β Gal; blue, nuclei.

Cells expressing $N^{\Delta E(V1744)}$ mutants have the same overall levels of Notch-1 protein as do cells expressing $N^{\Delta E}$ but differ in the amount of NICD produced and the degree of HES-1 activation (Figs 3a and 4a). This could result either from lower HES-1 reporter activation in the same number of cells, or from maximal activation of the HES-1 reporter in fewer cells. At saturating DNA concentrations, half as many cells transfected with $N^{\Delta E(V1744)}$ mutants as cells transfected with $N^{\Delta E}$ express the HES-1- β -galactosidase reporter (yellow and green bars, Fig. 5b). These data support the proposal that an increase in NICD concentration increases the probability of forming fully functional transcription complexes at the HES-1 promoter, rather than increasing the rate of transcription from a fixed number of templates²⁶. The mutations decrease this probability by decreasing the amounts of NICD produced.

The processing model predicts that ligand binding of full-length Notch-1 should result in NICD release. Such cleavage has not been observed previously by western blot analysis or immunohistochemistry (Fig. 6a). As immunoprecipitation of CSL^{RBP3} enriches for NICD, we used this technique to test whether the Notch-1 ligand Jagged-1²⁷ can induce cleavage of N^{FL} (see Methods) in tissue culture cells. In all experiments, a cleaved product that migrates together with NICD is upregulated when N^{FL} is transfected with the ligand Jagged-1 (Fig. 6b). To confirm the identity of this cleaved product as NICD, N^{FL} proteins containing cleavage mutations ($N^{FL(V1744L)}$ and $N^{FL(V1744K)}$) were also transfected with Jagged-1. In these experiments, no cleavage was observed (Fig. 6b). Thus, ligand binding can induce proteolytic processing of Notch-1 at V1744, releasing NICD.

In conclusion, activation of CSL by Notch-1 in vertebrate cells requires processing and nuclear localization of Notch-1, and V1744 is essential for efficient processing to occur. This same mechanism may also be vital in Notch-mediated oncogenesis²⁸. In addition, very low levels of NICD are enough to produce a significant response. Such low levels of NICD can be detected by co-precipitation with Flag- CSL^{RBP3} and, significantly, NICD can be found in cells following

interaction between Notch-1 and Jagged-1. Ligand-mediated processing of a membrane-bound transcription factor has also been demonstrated for the sterol regulatory-element-binding protein²⁹, which, unlike Notch, binds DNA directly. It is unknown whether processing is required for CSL-independent signalling events and whether processing of other Notch proteins occurs. □

Methods

Plasmids. Proteins were expressed using the vector pCS2³⁰. $N^{\Delta E13}$, N^{LNG13} , pCS2+ β -galactosidase³⁰, HES-1-luc²⁷, Flag- CSL^{RBP3} and Jagged-1²⁷ have been described. $N^{\Delta E}$ has the methionine at position 1,727 replaced with a valine¹³. pCS2+ $N^{\Delta E(ARAM)}$ is an in-frame deletion of amino acids 1,751–1,778. pCS2+ N^{FL} contains the Notch-1 protein up to amino acid 2,813, fused to a hexameric Myc tag at the carboxy terminus. pCS2+ luc contains the *HindIII* to *SalI* luciferase-coding fragment of pGL2-basic ligated in pCS2+ cut by *HindIII/XhoI*.

We mutated the $N^{\Delta E}$ processing site by amplifying two overlapping polymerase chain reaction (PCR) products. The outside primers RR66 (5'-CCCAAGCTTGATTAGGTGAC-3') and M95 (5'-TCGAACATTGACATCCATGCA-3') were used with appropriate inside primers containing mutations to form primary PCR products. Secondary PCR products were purified, digested with *HindIII* and *MscI* and ligated into the same sites in $N^{\Delta E}$. Primers used for mutagenesis at V1744 were TGTGGGCTGTGGGnnnCTGCTGT CCCGCAA and TTGCGGGACAGCAGnnnCCCACAGCCACA, with nnn representing the mutagenized codons CTG (in V1744L) and AAA (in V1744K). A *NotI* fragment containing the mutations were then moved into pCS2+ N^{FL} to make pCS2+ $N^{FL(V1744L)}$ and pCS2+ $N^{FL(V1744K)}$. To create pCS2+ $N^{ICV1744}$ and its mutants, the primer CCAGGATCCACCATGnnnCTGCTGTCCCAGCAAGC was synthesized. This primer was used with M95 to generate a product that was digested with *BamHI* and *BclI*, and ligated into the same sites in $N^{\Delta E}$. GTG was used for the V1744 codon.

Transfection and reporter gene assays. 3T3 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% bovine calf serum. 10T1/2 and 293T cells were cultured in DMEM supplemented with 10% fetal bovine serum. Notch-1 activity was determined by measuring luciferase activity in 3T3 cells following transient transfection with HES-1-luc¹¹. Transfections were performed as described¹³ except that volumes were scaled down to produce 2 ml of medium in a 6-well dish. Each well was transfected with a total of 2 μ g DNA consisting of: 400 ng HES-1-luc, 40 ng pCS2+ β -galactosidase, 0.4–400 ng pCS2+ Notch-1, with the remaining DNA made up to 2 μ g with empty pCS2+ vector. 100 μ l lysate was used to determine luciferase activity as described⁶. 5 μ l lysate was used to determine β -galactosidase concentration and equalize results for transfection efficiency. Assays for β -galactosidase were performed using the chemiluminescent substrate Galacton (Tropix). Experiments were performed in duplicate and repeated four times. Maximal HES-1-luc activation varied between 20- and 50-fold. For immunofluorescent examination, cells were transfected as above except that HES-1- β -galactosidase was used as a reporter and pCS2+ luc was used for equalization. Fixed and permeabilized cells were incubated with the monoclonal anti-Myc antibody 9E10 and a rabbit polyclonal anti- β -galactosidase antibody (5prime3prime). BODIPY-conjugated goat anti-rabbit and Cy3-conjugated goat anti-mouse antibodies were used as secondary antibodies. Nuclei were visualized with Hoechst stain. For cell counting, one diameter (45 mm² area) was scanned on each plate. pCS2+ luc controls show that transfection efficiency varied by less than twofold (not shown).

Viral expression and sequencing. A Sindbis virus expression system was used to overexpress Notch-1 protein for microsequencing. The 1,873-base-pair *SphI/SnaBI* fragment of $N^{\Delta E}$ containing the coding sequence was ligated into the plasmid pSinRep5 using the *SphI* and *StuI* sites to make pSR/ $N^{\Delta E}$. Viral particles were produced using the defective helper RNA DHBB as described¹⁵ and used to infect ten 100-mm dishes of 10T1/2 cells. At 48 h after infection, Notch-1 was isolated and processed for microsequencing¹³.

Pulse chase and drug treatments. Subconfluent monolayers of 10T1/2 cells were infected with SR/ $N^{\Delta E}$ pseudovirions and used 24 h after infection. Brefeldin A was prepared as a 1 mg ml⁻¹ stock in methanol and used at a concentration of 2 μ g ml⁻¹. Monensin was prepared as a 10 mM stock in ethanol and used at a concentration of 10 μ M. Labelling experiments and

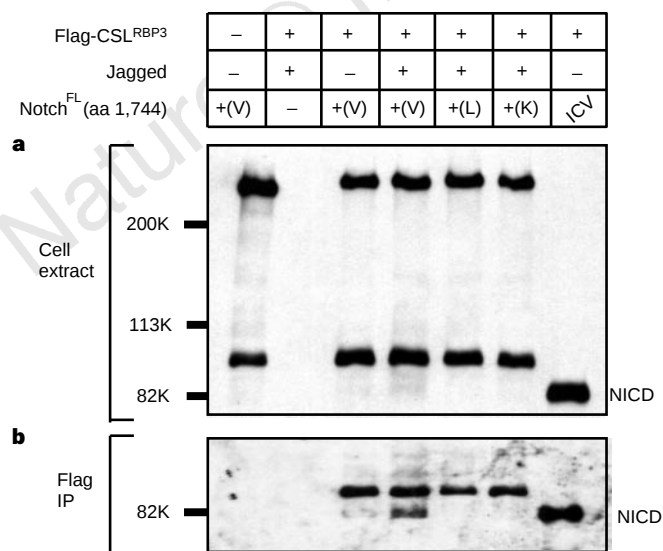


Figure 6 Full-length Notch-1 is cleaved when expressed together with Jagged-1. N^{FL} , CSL^{RBP3} and Jagged-1 were all expressed in kidney epithelial 293 cells, which allow high transfection efficiency and cell-cell contact. *In vivo*, ligand and receptor are often expressed in the same cell, but Notch activation results from ligand presented by one cell to Notch on another cell¹. Notch-1 was co-precipitated as in Fig. 3. NICD is markedly upregulated when N^{FL} is transfected with Jagged-1 and is absent in the processing mutants $N^{FL(V1744L)}$ and $N^{FL(V1744K)}$. Transfection of full-length Notch-1 and CSL^{RBP3} alone does not generate similar fragments. **a**, Whole-cell extract. **b**, Immunoprecipitation (IP) with Flag. + (V), + (L), + (K) and the V of ICV indicate the amino acid at position 1,744 of N^{FL} .

immunoprecipitations were done as described¹³. Ethanol and methanol controls did not alter processing.

Co-precipitations. Notch-1 proteins were precipitated together with Flag-CSL^{RBP3} from transiently transfected 293T cells as described⁷, with the following modifications: anti-Flag antibody M2 (Kodak) was used to precipitate Flag-CSL^{RBP3} and Myc-tagged Notch-1 proteins were detected with monoclonal antibody 9E10. Duplicate 60-mm plates were transfected with 1.5 µg Notch-1 plasmid and/or pCS2⁺ Jagged-1 and pCS2⁺ Flag-CSL^{RBP3} (3 µg), and pCS2⁺ vector was added to a total of 6 µg. Cells were lysed and pooled in 1 ml buffer⁷. Notch-1 does not precipitate from cell extracts that were transfected separately with CSL^{RBP3} and Notch-1 and then mixed (data not shown).

Received 22 December 1997; accepted 22 April 1998.

- Artavanis-Tsakonas, S., Matsuno, K. & Fortini, M. E. Notch signaling. *Science* **268**, 225–232 (1995).
- Kopan, R. & Turner, D. The Notch pathway: democracy and aristocracy during the selection of cell fate. *Curr. Opin. Neurobiol.* **6**, 594–601 (1996).
- Weinmaster, G. The ins and outs of Notch signaling. *Mol. Cell. Neurosci.* **9**, 91–102 (1997).
- Artavanis-Tsakonas, S. Alagille syndrome—a notch up for the Notch receptor. *Nature Genet.* **16**, 212–213 (1997).
- Fortini, M. E. & Artavanis-Tsakonas, S. The suppressor of hairless protein participates in Notch receptor signaling. *Cell* **79**, 273–282 (1994).
- Roehl, H., Bosenberg, M., Belloch, R. & Kimble, J. Roles of the Ram and Ank domains in signaling by the *C. elegans* Glp-1 receptor. *EMBO J.* **15**, 7002–7012 (1996).
- Aster, J. C. *et al.* Oncogenic forms of Notch1 lacking either the primary binding site for Rbp-J-kappa or nuclear localization sequences retain the ability to associate with Rbp-J-kappa and activate transcription. *J. Biol. Chem.* **272**, 11336–11343 (1997).
- Lieber, T., Kidd, S., Alcamo, E., Corbin, V. & Young, M. W. Antineurogenic phenotypes induced by truncated Notch proteins indicate a role in signal transduction and may point to a novel function for Notch in nuclei. *Genes Dev.* **7**, 1949–1965 (1993).
- Kopan, R., Nye, J. S. & Weintraub, H. The intracellular domain of mouse Notch: a constitutively activated repressor of myogenesis directed at the basic helix-loop-helix region of MyoD. *Development* **120**, 2385–2396 (1994).
- Hsieh, J. J.-D. & Hayward, S. D. Masking of the CBF1/RBPjk transcriptional repression domain by Epstein-Barr virus EBNA2. *Science* **268**, 560–563 (1995).
- Jarriault, S. *et al.* Signalling downstream of activated mammalian Notch. *Nature* **377**, 355–358 (1995).
- Tamura, K. *et al.* Physical interaction between a novel domain of the receptor Notch and the transcription factor Rbp-J-kappa/Su(H). *Curr. Biol.* **5**, 1416–1423 (1995).
- Kopan, R., Schroeter, E. H., Weintraub, H. & Nye, J. S. Signal transduction by activated mNotch: importance of proteolytic processing and its regulation by the extracellular domain. *Proc. Natl Acad. Sci. USA* **93**, 1683–1687 (1996).
- Wettstein, D. A., Turner, D. L. & Kintner, C. The *Xenopus* homolog of *Drosophila* Suppressor Of Hairless mediates Notch signaling during primary neurogenesis. *Development* **124**, 693–702 (1997).
- Bredenoek, P. J., Frolov, I., Rice, C. M. & Schlesinger, S. Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. *J. Virol.* **67**, 6439–6446 (1993).
- Fujiwara, T., Oda, K., Yokota, S., Takatsuki, A. & Ikebara, Y. Brefeldin A causes disassembly of the Golgi complex and accumulation of secretory proteins in the endoplasmic reticulum. *J. Biol. Chem.* **263**, 18545–18552 (1988).
- Doms, R. W., Russ, G. & Yewdell, J. W. Brefeldin A redistributes residence and itinerant Golgi proteins to the endoplasmic reticulum. *J. Cell Biol.* **109**, 61–72 (1989).
- Griffiths, G., Quinn, P. & Warren, G. Dissection of the Golgi complex. I. Monensin inhibits the transport of viral membrane proteins from medial to trans Golgi cisternae in baby hamster kidney cells infected with Semliki Forest virus. *J. Cell Biol.* **96**, 835–850 (1983).
- Quinn, P., Griffiths, G. & Warren, G. Dissection of the Golgi complex. II. Density separation of specific Golgi functions in virally infected cells treated with monensin. *J. Cell Biol.* **96**, 851–856 (1983).
- Pan, D. J. & Rubin, G. M. Kuzbanian controls proteolytic processing of Notch and mediates lateral inhibition during *Drosophila* and vertebrate neurogenesis. *Cell* **90**, 271–280 (1997).
- Sotillos, S., Roch, F. & Campuzano, S. The metalloprotease-disintegrin Kuzbanian participates in Notch activation during growth and patterning of *Drosophila* imaginal discs. *Development* **124**, 4769–4779 (1997).
- Wen, C., Metzstein, M. M. & Greenwald, I. Sup-17, a *Caenorhabditis elegans* ADAM protein related to *Drosophila* KUZBANIAN, and its role in the Lin-12/Notch signaling. *Development* **124**, 4759–4767 (1997).
- Blaumueller, C. M., Qi, H. L., Zagouras, P. & Artavanis-Tsakonas, S. Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane. *Cell* **90**, 281–291 (1997).
- Gho, M., Lecourtis, M., Geraud, G., Posakony, J. W. & Schweisguth, F. Subcellular localization of Suppressor of Hairless in *Drosophila* sense organ cells during Notch signalling. *Development* **122**, 1673–1682 (1996).
- Honjo, T. The shortest path from the surface to the nucleus: RBPjk/Su(H) transcription factor. *Genes Cells* **1**, 1–9 (1996).
- Weintraub, H. Formation of stable transcription complexes as assayed by analysis of individual templates. *Proc. Natl Acad. Sci. USA* **85**, 5819–5823 (1988).
- Lindsell, C. E., Shawber, C. J., Boulter, J. & Weinmaster, G. Jagged: a mammalian ligand that activates Notch1. *Cell* **80**, 909–917 (1995).
- Capobianco, A. J., Zagouras, P., Blaumueller, C. M., Artavanis-Tsakonas, S. & Bishop, J. M. Neoplastic transformation by truncated alleles of human NOTCH1/TAN1 and NOTCH2. *Mol. Cell. Biol.* **17**, 6265–6273 (1997).
- Brown, M. S. & Goldstein, J. L. The SREBP pathway—regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
- Turner, D. L. & Weintraub, H. Expression of achaete-scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8**, 1434–1447 (1994).

Acknowledgements. We thank A. Israel for providing us with the FLAG-CSL^{RBP3} expression vector; I. Frolov for the Sindbis virus vectors pSINrep5 and DHBB; G. Rubin and D. J. Pan for the Kuzbanian cDNAs; D. Syder and M. Crankshaw for technical help; R. Cagan, I. Boime, J. Gordon, D. Ornitz and members of the Kopan lab for reading and commenting on the manuscript; and G. Goldberg, G. Grant and B. Marmer for discussions and technical advice. R. K. Thanks H. Weintraub for encouragement and support. This research was supported by the NIH.

Correspondence and requests for materials should be addressed to R.K. (e-mail: kopan@pharmdec.wustl.edu).

Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex

Xinsheng Nan^{††}, Huck-Hui Ng^{††}, Colin A. Johnson[‡], Carol D. Laherty[§], Bryan M. Turner[‡], Robert N. Eisenman[§] & Adrian Bird^{*}

^{*} Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JR, UK

[‡] Department of Anatomy, The Medical School, University of Birmingham, Birmingham B15 2TT, UK

[§] Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, USA

^{††} These authors contributed equally to this work.

Cytosine residues in the sequence 5' CpG (cytosine–guanine) are often postsynthetically methylated in animal genomes. CpG methylation is involved in long-term silencing of certain genes during mammalian development^{1,2} and in repression of viral genomes^{3,4}. The methyl-CpG-binding proteins MeCP1 (ref. 5) and MeCP2 (ref. 6) interact specifically with methylated DNA and mediate transcriptional repression^{7–9}. Here we study the mechanism of repression by MeCP2, an abundant nuclear protein that is essential for mouse embryogenesis¹⁰. MeCP2 binds tightly to chromosomes in a methylation-dependent manner^{11,12}. It contains a transcriptional-repression domain (TRD) that can

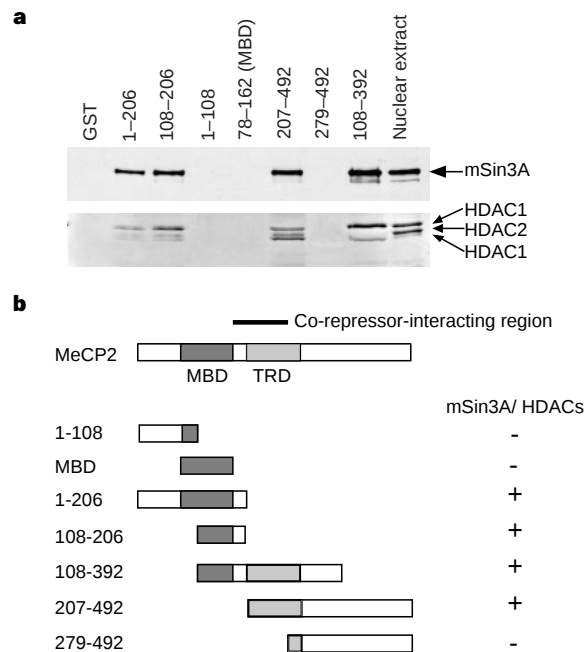


Figure 1 Interaction of MeCP2 with mSin3A and histone deacetylases. **a**, Western blot analysis of HeLa nuclear proteins pulled down by immobilized GST or by GST-fusion proteins that included different regions of MeCP2 (see **b**). Numbers correspond to amino-acid positions in the protein. Blots were probed with antibodies against mSin3A (antibody K-20), or against HDAC1 and HDAC2. The lower HDAC1 band probably corresponds to a degradation product of HDAC1. The HDAC2 band in lane 108–392 is displaced upwards because a large amount of fusion protein migrates with the HDAC2. **b**, Map of regions of MeCP2 that do (+) or do not (-) pull down mSin3A and HDACs. The thick black line corresponds to the corepressor-interacting region defined by these experiments. MBD, methyl-CpG-binding domain¹¹.

function at a distance *in vitro* and *in vivo*⁹. We show that a region of MeCP2 that localizes with the TRD associates with a corepressor complex containing the transcriptional repressor mSin3A and histone deacetylases^{13–19}. Transcriptional repression *in vivo* is relieved by the deacetylase inhibitor trichostatin A²⁰, indicating that deacetylation of histones (and/or of other proteins) is an essential component of this repression mechanism. The data suggest that two global mechanisms of gene regulation, DNA methylation and histone deacetylation, can be linked by MeCP2.

DNA methylation is associated with altered chromatin structures^{21–23}. As MeCP2 binds tightly to chromatin in a methylation-dependent manner, it seemed possible that transcriptional repression may be due to recruitment by MeCP2 of a chromatin-modifying corepressor. A corepressor complex containing mSin3 and histone deacetylases (HDACs) is thought to exert its effects by modulating chromatin structure^{13–19}. We therefore asked whether regions of MeCP2 could 'pull down' the mSin3-containing corepressor complex from nuclear extracts. We used proteins containing glutathione S-transferase (GST) fused to different regions of MeCP2 in 'pull-down' reactions, and probed proteins bound to these complexes with antibodies against mSin3A, HDAC1 and HDAC2. A region in the centre of MeCP2 could bind all three proteins (Fig. 1a). GST alone and regions outside the interacting region of MeCP2 did not pull down any of the corepressor proteins (Fig. 1a).

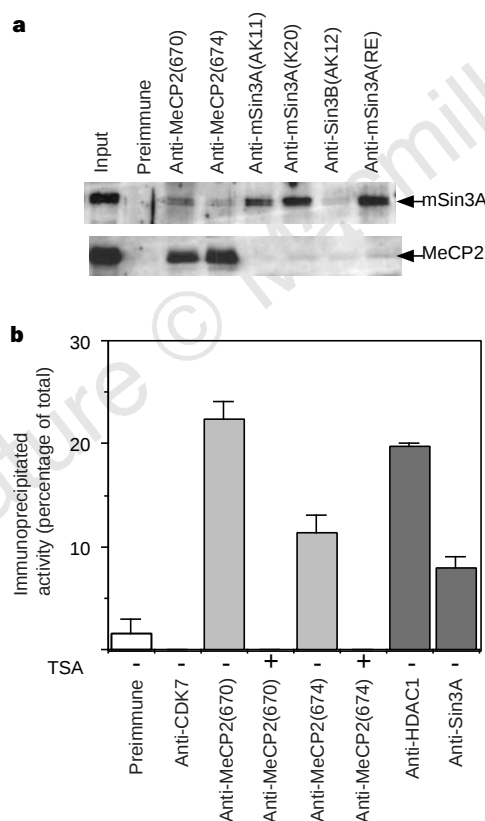


Figure 2 Co-immunoprecipitation of mSin3, MeCP2 and histone deacetylase activity from rat brain nuclear extracts. **a**, Western blot of immunoprecipitates prepared using antibodies against MeCP2 (antibodies 670 and 674) or mSin3 (antibodies AK11, AK12, K-20 and RE) or using preimmune serum. Probes were anti-mSin3A (RE) and anti-MeCP2 (670 and 674) antibodies. **b**, Immunoprecipitation of histone deacetylase activity from rat brain nuclear extracts with antibodies against MeCP2 (670 and 674), HDAC1, mSin3A or cyclin-dependent kinase 7 (CDK7) or with preimmune serum, in the presence or absence of 5 ng ml⁻¹ TSA. Precipitated activities from duplicate parallel experiments are shown as percentage total activity added to each tube (19,200 d.p.m. per 2 h in the experiment shown).

Control nuclear proteins (proliferating cell nuclear antigen and TATA-binding protein) were not bound by MeCP2 (data not shown). The region of MeCP2 that interacted with the corepressor substantially overlaps the TRD, as defined *in vivo* (Fig. 1b)⁹.

To test for a possible *in vivo* interaction between native MeCP2 and the mSin3 corepressor complex, we used antibodies against MeCP2 and mSin3 to immunoprecipitate proteins from rat brain nuclear extracts (Fig. 2a). Both anti-MeCP2 antibodies precipitated mSin3A as well as MeCP2. Four different anti-mSin3 antibodies immunoprecipitated MeCP2 as well as mSin3 (the anti-Sin3 antibody AK11 precipitated less MeCP2 than the other antibodies). Immunoprecipitates of transiently expressed HDAC2 also contained MeCP2 (data not shown). These results led us to expect that MeCP2 would be associated with a catalytically active deacetylase complex. This was confirmed by the finding that both anti-MeCP2 antibodies precipitated deacetylation activity, which was abolished by the highly specific deacetylase inhibitor trichostatin A (TSA)²⁰ (Fig. 2b). Antibodies against components of the corepressor complex, mSin3A and HDAC1, also immunoprecipitated deacetylase activity, as expected, whereas control antibodies (anti-CDK7 antibody and preimmune serum) gave background activity (Fig. 2b).

Does MeCP2 interact with mSin3 itself or with another component of the complex? To answer this question, we assayed the ability

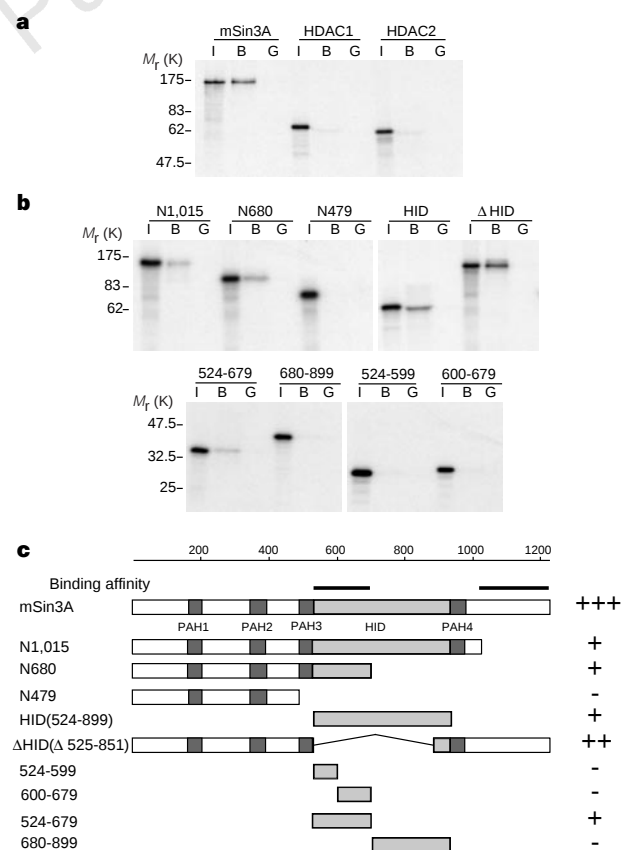


Figure 3 MeCP2 interacts with mSin3A, but not with HDACs. **a**, *In vitro*-translated ³⁵S-labelled mSin3A and HDACs 1 and 2 were incubated with an immobilized fusion protein containing GST and amino acids 108–392 of MeCP2 (Fig. 1b). Input protein (I) (20% of total), protein bound by the 108–392 fusion protein (B) and protein bound by GST alone (G) were analysed by SDS-PAGE. **b**, Deletion constructs of mSin3A were transcribed and translated *in vitro* to give ³⁵S-labelled products and processed as in **a**. Input (I) corresponds to 10% of total. **c**, Maps of the mSin3A deletion constructs¹³, indicating truncated proteins that did (+) or did not (-) bind to MeCP2. Regions of strong binding to MeCP2 are indicated by the thick black lines (top).

of the corepressor-interacting region of MeCP2 (residues 108–392) to bind *in vitro*-translated mSin3A, HDAC1 and HDAC2. mSin3A is the preferred binding partner, as the HDACs have a much weaker affinity for MeCP2 (Fig. 3a). There is probably a direct interaction between MeCP2 and mSin3A, although we cannot exclude the possibility that an unknown component of the translation lysate mediates indirect binding. We mapped the sites of interaction between MeCP2 and mSin3A by translating *in vitro* a series of mSin3A proteins with deleted portions (ref. 13) and testing their ability to bind to amino acids 108–392 of MeCP2 (Fig. 3b, c). The amino-terminal half of the HDAC-interaction domain (HID)¹³ of mSin3A was common to most fragments that were able to binding MeCP2, but a deletion lacking all but the carboxy-terminal 48 amino acids of HID (Δ HID) also bound efficiently, indicating a strong binding site within the C-terminal 200 amino acids of mSin3A (Fig. 3). We infer that the interaction between MeCP2 and mSin3A involves both the N-terminal half of the HID of mSin3A and the C-terminus of mSin3A. MeCP2 and deacetylases seem to bind to distinct sites within the 375-residue HID region, as both can bind simultaneously (Figs 1a and 2a, b).

To determine whether deacetylation is important for the effects of the TRD on transcription *in vivo*, we attempted to alleviate

transcriptional repression with TSA²⁰. We transiently transfected mouse L929 fibroblasts with a reporter gene, which either did (G₅) or did not (G₀) have GAL4-binding sites near its promoter, in the presence or absence of a GAL4–TRD fusion construct⁹. In the absence of repressor, the ratio of G₅ expression to control G₀ expression was unaffected by TSA, remaining close to unity (Fig. 4). In the presence of the GAL4–TRD repressor, the G₅:G₀ expression ratio was about 0.2, but was increased 2.8-fold by TSA. Incubations with TSA for 24 h (Fig. 4) or 9 h (data not shown) gave essentially the same result, eliminating the possibility that TSA exerted its effects by arresting the cells at a stage of the cell cycle where repression was minimal. Repression by GAL4–TRD was not completely alleviated by TSA in any of the experiments (Fig. 4, data not shown). The results indicate that a component of repression by the TRD may be deacetylase-independent, consistent with the observation that mSin3A retains some ability to repress transcription even in the absence of associated HDACs¹³.

Our results indicate that repression by the TRD of MeCP2 relies, to a significant extent, on histone deacetylation. Evidence for a functional relationship between DNA methylation and chromatin has come from studies showing that methylated DNA packaged as chromatin is incapable of transcription, whereas methylated DNA alone or chromatin alone remains transcriptionally active^{22,23}. Given the strong correlation between deacetylation of histone protein tails and transcriptional repression²⁴, the idea that methylated DNA provokes deacetylation is attractive²⁵, and is supported by evidence that TSA can substitute for the demethylating drug 5-azacytidine to derepress methylated ribosomal RNA genes in plants²⁶. Our results indicate that MeCP2 may provide a mechanistic bridge between DNA methylation and histone deacetylation. Recruitment of the corepressor complex by chromatin-bound MeCP2 may lead to local deacetylation of core histones (and perhaps of other proteins involved in transcription²⁷), with consequent elimination of transcription. □

Methods

GST pulldown assay. GST and GST fusion proteins¹¹ (6 μ g) were first immobilized on glutathione–Sepharose beads. The coated beads were then incubated with 60 μ g HeLa cell nuclear extract²⁸ in buffer A (20 mM HEPES pH 7.9, 150 mM NaCl, 0.5 mM EDTA, 1 mM dithiothreitol, 10% glycerol, 0.1% Triton X-100) for 2 h at 4 °C. After five washes with buffer A, the proteins bound to the beads were eluted in Laemmli buffer and analysed by western blotting nearly as described¹³. The primary antibodies used were K20 (anti-mSin3A, Santa Cruz Biotech. Inc.), anti-HDAC1, and anti-HDAC2 (ref. 13). Nuclear extract (6 μ g protein) was loaded as a positive control. Polyclonal antiserum against mammalian HDAC1 was raised against a synthetic peptide corresponding to the C-terminal domain of HDAC1 (EEKPEAKGVKEEVKLA) as described²⁹.

Immunoprecipitation. Antibodies (4 μ l serum, or 2 μ g purified IgG) were immobilized on protein A–Sepharose beads and washed with 20 mM HEPES pH 7.9, 0.1 M NaCl, 10% glycerol, 0.2 mM EDTA and 0.01% Triton X-100. The antisera were: anti-MeCP2 (670) raised against the amino acids 207–492 of MeCP2; anti-MeCP2 (674) raised against amino acids 1–392 of MeCP2; anti-mSin3A (RE)¹³; and anti-mSin3A/B antibodies AK-11, K-20 and AK12 (Santa Cruz Biotech. Inc.) Preimmune serum was from the rabbit that was subsequently immunized with MeCP2 amino acids 207–492. Bound antibodies were incubated with 20 μ g brain nuclear extract^{6,28} in 240 μ l wash buffer plus 1 mg ml⁻¹ BSA for 1.5 h. Beads were washed three times and protein was eluted in Laemmli buffer. Immunoprecipitated proteins were detected by western blotting.

Immunoprecipitation of histone deacetylase activity. Aliquots (100 μ g protein) of rat brain nuclear extract were incubated for 16 h at 4 °C, with 5 μ g affinity-purified antibodies against HDAC1 or mSin3A (Santa Cruz Biotech. Inc.) or with whole antiserum against MeCP2 (670 or 674) or CDK7 (a gift from J. Shuttleworth) or with preimmune serum. The final volume was 500 μ l in 20 mM HEPES pH 7.9, 150 mM NaCl, 0.1% Nonidet P40, 0.25% gelatin, 1 mM EDTA, 0.5 mM EGTA, 0.2% NaN₃, 0.1 mM PMSF, and 1 μ g ml⁻¹ each of

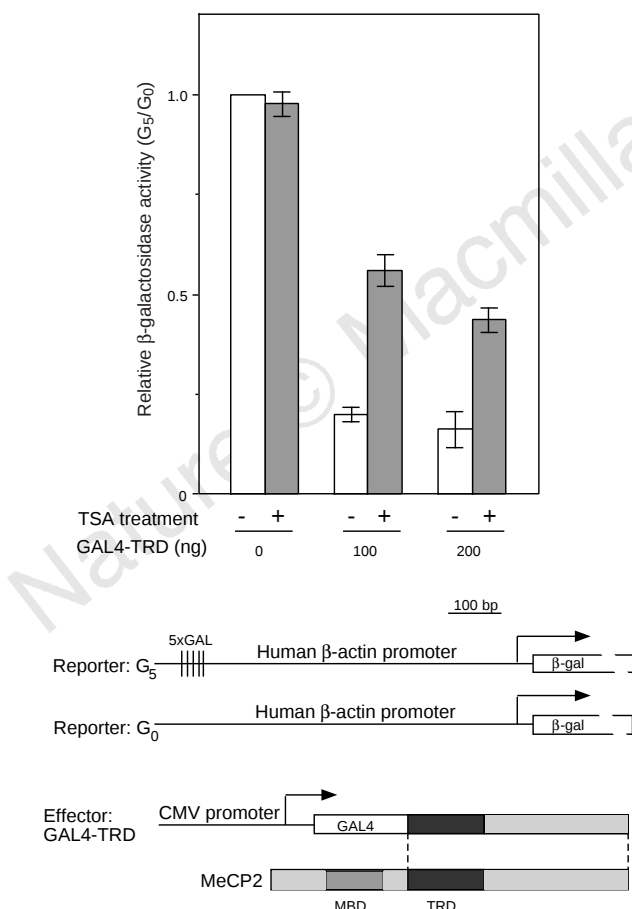


Figure 4 Relief of TRD-mediated repression of transcription by the deacetylase inhibitor TSA. Mouse fibroblasts (L cells) were transfected with reporter genes that did or did not contain GAL4 DNA-binding sites (G₅ or G₀, respectively), in the presence or absence of a fusion between the GAL4 DNA-binding domain and amino acids 207–492 of MeCP2. After 24 h, transfected cells were treated with TSA (shaded bars) or left untreated (white bars) and were incubated for a further 24 h. Columns represent the mean ratio between β -galactosidase activities of G₅ and G₀ in three independent experiments ($\pm$ standard deviation). Relative β -galactosidase activity in the absence of TSA and GAL4–TRD was normalized to 1.0. CMV, cytomegalovirus; bp, base pairs.

aprotinin, leupeptin and pepstatin A. Antibody-bound material was pelleted with protein A-Sepharose (Pharmacia), washed and assayed for histone deacetylase activity as described³⁰. The substrate was a synthetic peptide corresponding to the 18 N-terminal residues of histone H4, chemically acetylated using ³H-labelled acetic anhydride (Amersham) and added at 3 × 10⁵ d.p.m. per assay.

In vitro binding assay. Intact mSin3A, HDAC1 (a gift from T. Kouzarides), HDAC2 and fragments of mSin3A (ref. 13) were labelled with [³⁵S]methionine using a coupled transcription-translation TNT system (Promega). GST-MeCP2 fusion proteins or GST (3 µg) were first bound to glutathione-Sepharose beads. The slurry was then incubated with the labelled products (1 µl) in 100 µl buffer B (50 mM HEPES pH 7.9, 250 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 10% glycerol, 0.1% Triton X-100, 0.5 mg ml⁻¹ BSA) at 4 °C for 2 h. After washing the beads for three times with buffer B, the bound proteins were eluted in Laemmli loading buffer, fractionated by SDS-PAGE, and detected by autoradiography.

Transfection and β-galactosidase assay. Mouse fibroblasts (L cells) were transfected with reporters G₀ or G₅ in the presence or absence of GAL4-TRD(207–492). Transfection and β-galactosidase assays were performed nearly as described⁹. Fresh medium with or without 100 ng ml⁻¹ TSA was added to the mouse L929 fibroblast cells 24 h after transfection and the cells were collected after a further 24 h. In some experiments, treatment with TSA was reduced to 9 h.

Received 9 February; accepted 19 March 1998.

- Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
- Beard, C., Li, E. & Jaenisch, R. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes Dev.* **9**, 2325–2334 (1995).
- Jahner, D. *et al.* De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623–628 (1982).
- Groudine, M., Eisenman, R. & Weintraub, H. Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. *Nature* **292**, 311–317 (1981).
- Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499–507 (1989).
- Lewis, J. D. *et al.* Purification, sequence and cellular localisation of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905–914 (1992).
- Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123–1134 (1991).
- Cross, S. H., Meehan, R. R., Nan, X. & Bird, A. A component of the transcriptional repressor MeCP1 is related to mammalian DNA methyltransferase and thiorax-like protein. *Nature Genet.* **16**, 256–259 (1997).
- Nan, X., Campoy, J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471–481 (1997).
- Tate, P., Skarnes, W. & Bird, A. The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nature Genet.* **12**, 205–208 (1996).
- Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **21**, 4886–4892 (1993).
- Nan, X., Tate, P., Li, E. & Bird, A. P. DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.* **16**, 414–421 (1996).
- Laherty, C. D. *et al.* Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* **89**, 849–856 (1997).
- Alland, L. *et al.* Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature* **387**, 49–55 (1997).
- Hassig, C. A., Fleischer, T. C., Billin, A. N., Schreiber, S. L. & Ayer, D. E. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **89**, 341–347 (1997).
- Nagy, L. *et al.* Nuclear receptor repression mediated by a complex containing SMRT, mSINA, and histone deacetylase. *Cell* **89**, 373–380 (1997).
- Heinzel, T. *et al.* A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* **387**, 43–48 (1997).
- Zhang, Y., Iatni, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89**, 357–364 (1997).
- Kadosh, D. & Struhl, K. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* **89**, 365–371 (1997).
- Yoshida, M., Horinouchi, S. & Beppu, T. Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *Bioessays* **17**, 423–430 (1995).
- Lewis, J. & Bird, A. DNA methylation and chromatin structure. *FEBS Lett.* **285**, 155–159 (1991).
- Kass, S. U., Landsberger, N. & Wolffe, A. P. DNA methylation directs a time-dependent repression of transcription initiation. *Curr. Biol.* **7**, 157–165 (1997).
- Buschhausen, G., Witting, B., Graessmann, M. & Graessmann, A. Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc. Natl Acad. Sci. USA* **84**, 1177–1181 (1987).
- Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
- Kass, S. U., Pruss, D. & Wolffe, A. P. How does DNA methylation repress transcription? *Trends Genet.* **13**, 444–449 (1997).
- Chen, Z. J. & Pikaard, C. S. Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation histone modification in nucleolar dominance. *Genes Dev.* **11**, 2124–2136 (1997).
- Pazin, M. J. & Kadonaga, J. T. What's up and down with histone deacetylation and transcription? *Cell* **89**, 325–328 (1997).
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **5**, 1475–1489 (1983).
- Turner, B. M. & Fellows, G. Specific antibodies reveal ordered and cell cycle-related use of histone H4 acetylation sites in mammalian cells. *Eur. J. Biochem.* **179**, 131–139 (1989).

- Li, W., Chen, H. Y. & Davie, J. R. Properties of chicken erythrocyte histone deacetylase associated with the nuclear matrix. *Biochem. J.* **314**, 631–637 (1996).

Acknowledgements. We thank J. Davidson and A. Greig for technical assistance, E. Seto for anti-HDAC2 antiserum and T. Kouzarides for the HDAC1 expression plasmid. The work was supported by a grant from the N.I.H. (to R.N.E.) and by grants from The Wellcome Trust (to A.B. and B.M.T.). H.-H.N. is a Darwin Trust scholar; C.D.L. is a Leukemia Society of America Special Fellow; and R.N.E. is an American Cancer Society Research Professor.

Correspondence and requests for materials should be addressed to A.B. (e-mail: A.Bird@ed.ac.uk).

Transcriptional activation independent of TFIID kinase and the RNA polymerase II mediator *in vivo*

Dong-ki Lee & John T. Lis

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853, USA

The carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II becomes multiply phosphorylated by protein kinases during early steps in the gene transcription cycle both *in vivo*¹ and *in vitro*². In yeast, the major CTD kinase is a subunit of the general transcription factor TFIID, and is encoded by an essential gene, *KIN28* (ref. 3). Although the CTD and its phosphorylation are important for transcription^{4–6}, *in vitro* studies^{7,8} have challenged whether CTD phosphorylation is an absolutely required step. The general importance of CTD phosphorylation by Kin28 for transcription in yeast has been suggested because, for all genes tested, transcription is inhibited at the non-permissive temperature in temperature-sensitive *kin28* mutants^{9,10}. However, using such a mutant and a copper-inducible targeted destruction method, we show here that transcription of certain genes can be highly induced even when cells lack Kin28. We also show that transcription of these Kin28-independent genes is independent of Srb4 and Srb6, critical components of the CTD-associated transcriptional mediator complex¹¹. These results indicate that there are at least two distinct pathways for transcriptional activation: one is dependent on Kin28 and the mediator complex, and the other is not.

In agreement with previous studies^{9,10}, constitutive transcription of the *MET19* and *ADH1* genes (Fig. 1a) and inducible transcription of the *GAL10* gene (Fig. 1b, left panel) were severely affected at the non-permissive temperature in a yeast (*Saccharomyces cerevisiae*) strain containing a conditional mutation in the TFIID kinase subunit (*kin28-ts* mutant). Although some *ADH1* messenger RNA is still present at 1 h after the temperature shift because of its long half-life (>30 min), a run-on transcription experiment (Fig. 1a, right panel) shows that the level of polymerase found on the *ADH1* gene has decreased by >95%. We were surprised to find that some genes could be transcriptionally activated in cells that were severely depleted of this essential kinase and general transcription factor activity. After inactivating Kin28 by incubating the *kin28-ts* mutant at 37 °C for 1.5 h, the *CUP1* gene, a copper-inducible gene encoding metallothionein in yeast¹², remained highly inducible (85% of wild-type (WT) induction level) by copper sulphate (Fig. 1b, right panel). On the other hand, mutation in the largest subunit of RNA polymerase II (Pol II) (*rpb1-1*) severely decreased (to ~3% of WT levels) *CUP1* activation at the restrictive temperature (Fig. 1b, right panel). Thus, our data and those of others^{9,10} show that many genes require Kin28 for transcription; however, here we show *CUP1* activation by copper can occur in cells in which Kin28 has been inactivated.

TFIID contains several enzymatic activities, including ATPase

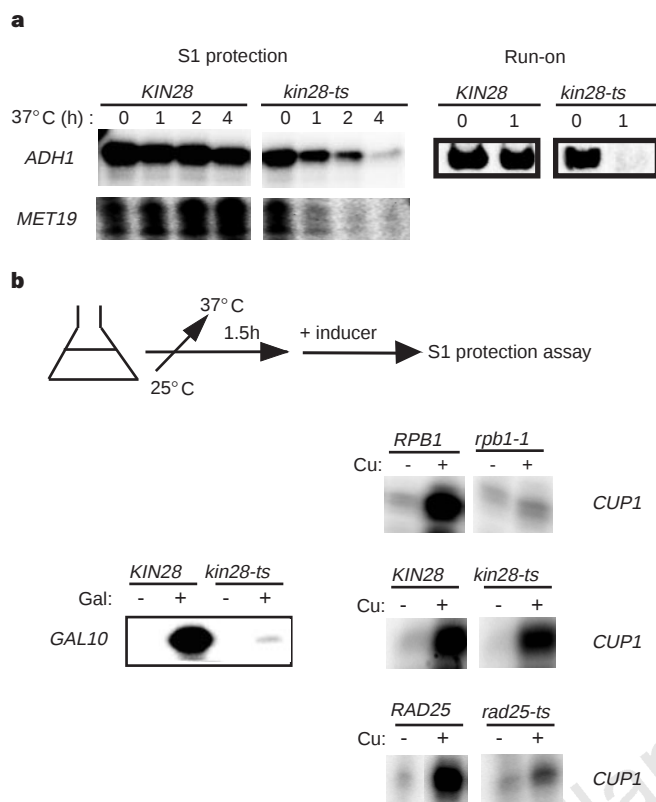


Figure 1 *CUP1* gene activation is independent of TFIH kinase, but dependent on TFIH helicase activity. **a**, Transcription of constitutive genes. A wild-type *S. cerevisiae* strain, GF262-2, and a *kin28-ts* strain, JGV4, were grown and samples were taken for RNA preparation and analysis before (0) and 1, 2, and 4 h after a shift to 37°C. *ADH1* and *MET19* RNAs were detected by S1 nuclease protection¹⁶ (left). To assess the polymerase density on the *ADH1* gene, run-on transcription was performed as described³⁰ (right) using the *KIN28* WT and ts strains after 0 and 1 h at the non-permissive temperature. Labelled RNAs were hybridized to a filter containing 5 µg of a PCR-amplified coding region of the *ADH1* gene and quantified using a Molecular Dynamics phosphorimager. **b**, Transcription of inducible genes. Cells were grown and incubated at 37°C for 1.5 h to pre-inactivate the activity of the relevant factor from the mutants shown. For *GAL10* gene induction, 2% galactose was added and cells were grown for a further 1 h. For *CUP1* gene induction, 1 mM CuSO₄ was added for 30 min. All mRNAs were detected by S1 nuclease protection and quantified. Serial dilutions of RNAs were performed to ensure the linearity of the assay. The quantified values in the text for given experiments are the average of two or three experiments where the variation between experiments is <20%.

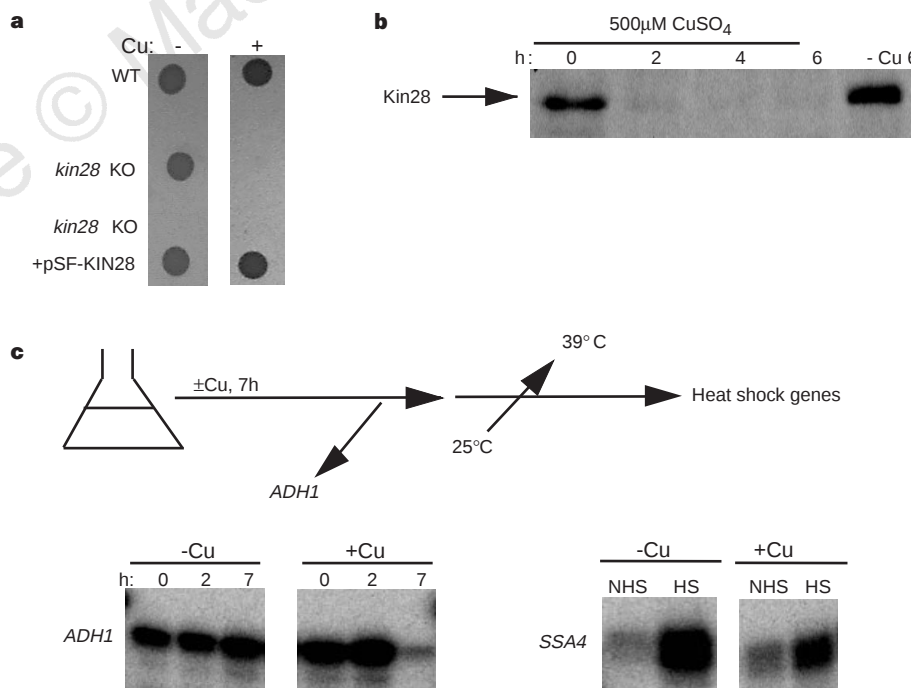


Figure 2 The *SSA4* gene is activated after depletion of Kin28. **a**, Colony growth of the copper-activated, double-shut-off strain (*kin28* KO) on a plate either lacking or containing 1 mM CuSO₄. KO, knockout. pSF-KIN28 encodes a wild-type *KIN28* gene. **b**, Western blot analysis after Kin28 depletion. Whole-cell extracts were prepared from yeast cultures grown in media either lacking or containing 1 mM CuSO₄. The extracts were probed with 12CA5 antibody to detect haemagglutinin-tag fused to Kin28, and visualized by chemiluminescence. **c**, *SSA4* gene

activation. The copper-activated double-shut-off strain was grown in media either lacking or containing 1 mM CuSO₄ for 7 h to deplete Kin28. Cells were then shifted to 39°C to induce transcription of the *SSA4* gene. *SSA4* RNA levels were measured by S1 nuclease protection. *ADH1* mRNA levels were also measured to show that transcription of a Kin28-dependent gene is inhibited on addition of copper. NHS, non heat-shock; HS, heat-shock.

and helicase activities, in addition to CTD-kinase activity. A temperature-sensitive mutant of *RAD25*, which encodes a helicase subunit, has been isolated¹³. General transcription was dramatically reduced in this mutant. We have tested whether the *CUP1* gene is also independent of this TFIIF helicase using a *rad25-ts* mutant. *CUP1* gene transcription is severely affected (reduced to ~3% of WT) at the restrictive temperature (Fig. 1b). Thus, although TFIIF kinase is dispensable for *CUP1* activation, at least this helicase component of core TFIIF is necessary.

Induction of yeast heat-shock genes is rapid and transient¹⁴, with peak accumulation of RNA at 10 min in the case of the heat-shock protein 70 gene, *SSA4* (ref. 15) (D.-k.L. and J.T.L., unpublished observations). As for *CUP1*, induction of *SSA4* is at least partially independent of Kin28 activity. As the temperature shift used to begin inactivation of *kin28-ts* rapidly activates transcription of heat-shock genes, we used the copper-inducible double-shut-off approach¹⁶ to generate a yeast strain that can be depleted of Kin28. The resulting conditional mutant, in which Kin28 is targeted for repression and degradation, did not grow on copper-containing plates (Fig. 2a). Introduction of a wild-type *KIN28* gene on a plasmid rescued this phenotype (Fig. 2a). Western blotting of whole-cell extracts from the conditional mutant showed the rapid depletion of ubiquitin-tagged Kin28 after the addition of copper (Fig. 2b). This strain was grown for 7 h with or without copper, and each culture was then heat-shocked. By this time, even levels of the constitutively expressed and rather stable *ADH1* RNA were dramatically reduced (to <10% of WT level) in the presence of copper (Fig. 2c, left panel). However, when we measured newly induced *SSA4* RNA levels in Kin28-depleted cells, we detected significant levels of RNA upon heat shock, albeit the levels were lower (~30% of WT level) than in the absence of copper (Fig. 2c, right panel). In contrast, the activation of the *SSA4* gene has been shown, using the same method, to be strictly dependent upon another general transcription factor, TFIIB¹⁶. Our result indicates that *SSA4* activation does occur when cellular Kin28 is reduced to a level that severely affects the transcription of most genes. Thus, it seems that Kin28 is not required for *SSA4* transcription, although it is possible that the *SSA4* gene can be transcribed when there are much lower intracellular Kin28 levels than are required for transcription of other genes. It has been suggested that heat shock decreases TFIIF kinase activity but activates another kinase to phosphorylate the transcribing Pol II in heat-stressed human cells¹⁷. Our results are the first direct demonstration that heat-shock genes can be activated without TFIIF kinase *in vivo*.

In yeast, the kinase CTK-1 was identified as a putative CTD kinase¹⁸. We tested whether CTK-1 is responsible for phosphorylation of the CTD in *CUP1* or *SSA4* gene transcription. Transcription of these genes in a *ctk-1*-deleted strain, however, was normal (data not shown). We also tested whether transcription of these genes requires the kinase *Srb10*, originally identified as a suppressor of a CTD truncation mutation¹⁹. The *CUP1* and *SSA4* genes were activated normally in an *srb10* mutant strain (data not shown). Therefore, at least three known yeast CTD kinases are dispensable for activation of *CUP1* and *SSA4*. These results indicate that CTD phosphorylation may not be necessary for transcription of these genes. However, we cannot rule out the possibility that another kinase, or perhaps several kinases, may be responsible for CTD phosphorylation in transcription of Kin28-independent genes.

Two models have been proposed to explain transcription activation in yeast^{20–23}. One model proposes that the upstream transcriptional activators interact with the CTD-bound mediator complex to recruit the holoenzyme form of Pol II to the promoter^{20,22,23}. The other proposes that transcriptional activators interact directly with components of TFIID to facilitate its recruitment to the promoter^{21,23}. These different models provide an explanation of the different sensitivities of genes after Kin28 inactivation. In the former model, Pol II holoenzyme is recruited by strong simulta-

neous interactions between activators and mediator²² and between mediator and CTD^{24,25}. One of these interactions needs to be severed for Pol II to clear the promoter and elongate the RNA being produced. CTD phosphorylation might be a way of allowing promoter clearance by disrupting the interaction between mediator and CTD, as the phosphorylated Pol II does not interact with the mediator²⁶. In the latter model, where TFIID recruitment is the rate-limiting step of the transcriptional activation, no interaction is obligatorily required between activator and mediator in the pre-initiation complex. As at least 50% of the Pol II in the cell is not associated with mediator^{24,25}, most Pol II molecules brought to the promoter by this mechanism may not contain mediator. Therefore, CTD phosphorylation may not be necessary to dissociate the CTD from the mediator to allow transcriptional activation by this mechanism.

One simple corollary of the latter model is that Kin28-independent genes would also be independent of the functional mediator. To test this, we have used *srb4-ts* and *srb6-ts* (temperature-sensitive) mutants to inactivate *Srb4* and *Srb6*, critical subunits of the mediator complex. In these mutants, transcription of all individual genes tested and most (>90%) Pol II transcription was abolished at the restrictive temperature¹¹. Here we show that transcription of these genes that are sensitive to Kin28 inactivation is also sensitive to *Srb4* inactivation (Fig. 3a). However, in the *srb4-ts* mutant, the

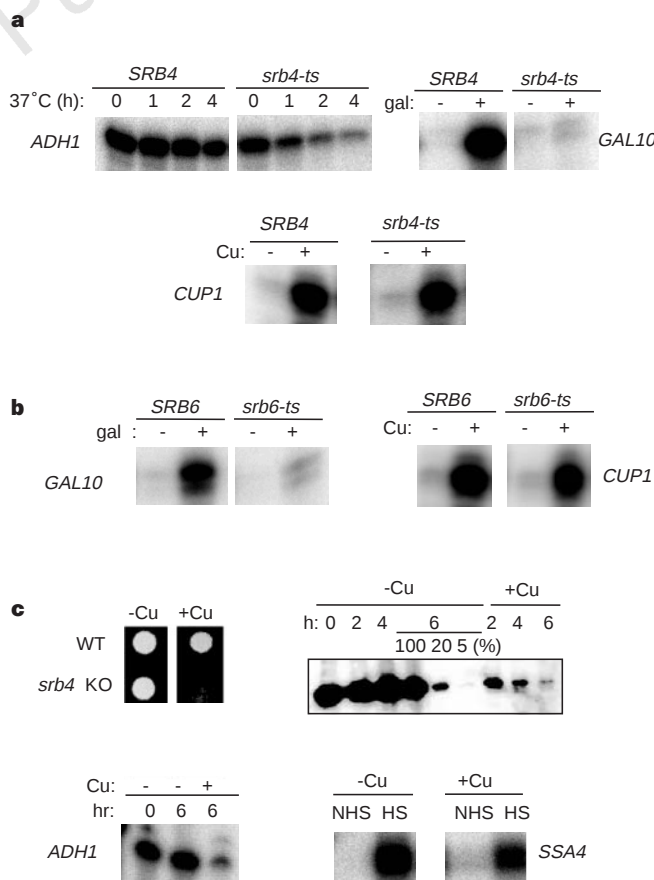


Figure 3 Kin28-independent genes are also independent of mediator. **a**, Transcription of genes in the *srb4-ts* mutant. Wild-type and *srb4-ts* strains were grown and shifted to 37°C; samples were then collected at the indicated times. *ADH1* mRNA was measured by S1 protection as in Fig. 1a. *GAL10* and *CUP1* induction experiments were performed as for Fig. 1b. **b**, Transcription of genes in the *srb6-ts* mutant. **c**, *SSA4* gene activation after *Srb4* depletion. Top left, colony growth of the copper-inducible double-shut-off *srb4*-knockout (KO) strain. Top right, western blot analysis of *Srb4* protein after *SRB4* deletion. Serial dilution (100%, 20%, 5%) of an extract collected at 6 h without copper was loaded. Bottom, analysis of *ADH1* and *SSA4* mRNA levels.

CUP1 gene can still be induced to ~60% of the WT level after pre-inactivating *Srb4* (Fig. 3a). The *CUP1* gene is fully induced after inactivating *Srb6* (Fig. 3b). To assess the effect of mediator depletion on *SSA4* gene activation, we generated a copper-inducible *Srb4*-knockout strain (Fig. 3c). After depleting *Srb4*, the *SSA4* gene is activated to a similar level (~30% of WT) as that seen in the *Kin28*-knockout strain (Fig. 3c). Therefore, a strong correlation exists between *Kin28*-independent and mediator-independent transcription.

We have found a set of genes that require neither TFIIF kinase nor normally critical components of the Pol II mediator complex for their activated transcription. From these findings and previous *in vitro*^{7,8} and *in vivo*⁵ data, we conclude that there are at least two distinct transcriptional activation mechanisms *in vivo*: one requires functional TFIIF kinase and mediator complex, and the other does not. Both of the activation mechanisms can be explained by the simple recruitment models described above. Interestingly, a second activity of TFIIF, the helicase activity, is required by both mechanisms. We have not addressed whether other mechanisms of transcriptional activation, such as those found in higher eukaryotes²⁷, may also exist in yeast. Nevertheless, our findings with TFIIF kinase and the mediator complex, with studies of promoter-specific TATA-binding protein (TBP)²⁸ and TBP-associated factors²⁹, show that transcription from various Pol II associated promoters *in vivo* is accomplished by different mechanisms, using different core transcriptional components. □

Methods

Strains and plasmids. *KIN28* WT and temperature-sensitive mutant strains were gifts from G. Faye¹⁰. *rpb1-1-ts* and *SRB4*, *SRB6* and *SRB10* WT and mutant strains were gifts from R. Young. *CTK1* WT and mutant strains were gifts from A. Greenleaf⁸. *RAD25* WT and temperature-sensitive strains were gifts from S. Prakash¹³. Copper-inducible knockout strains were generated as described¹⁶. Specifically, to generate *Kin28* KO, the *kin28* knockout strain, the 5' coding region of the *KIN28* gene, amplified by polymerase chain reaction, was cloned in-frame into *EcoRI*-*SacI* sites in plasmid ZM168. A *KpnI*-*SacI* fragment of this plasmid was moved to plasmid pRS306. The plasmid was digested with *XbaI* and transformed into the ZMY60 yeast strain. Colonies growing on Ura⁻ plates were scored for lethality on a plate containing 1 mM copper sulphate. To generate *Srb4* KO, the copper-inducible *srb4*-knockout strain, the 5' coding region of the *SRB4* gene was PCR-amplified and cloned into an *EcoRI*-*SacI* site in plasmid ZM168-HA3, a derivative of ZM168 that contains a three-haemagglutinin-epitope tag. From this plasmid, a *KpnI*-*SacI* fragment was moved to plasmid pRS306, and digested with *SpeI* to transform ZMY60. The pSF-KIN28 plasmid, a CEN/ARS plasmid containing an intact *KIN28* gene, was a gift from M. Cismowski and S. Reed⁹.

Received 6 November 1997; accepted 23 February 1998.

- O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J. T. Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature* **370**, 75–77 (1994).
- Laybourn, P. J. & Dahmus, M. E. Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* **265**, 13165–13173 (1990).
- Feaver, W. J., Svestrup, J. Q., Henry, N. L. & Kornberg, R. D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIFK. *Cell* **79**, 1103–1109 (1994).
- Scafe, C. et al. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature* **347**, 491–494 (1990).
- Gerber, H.-P. et al. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374**, 660–662 (1995).
- Akoulitchev, S., Makela, T. P., Weinberg, R. A. & Reinberg, D. Requirements for TFIIF kinase activity in transcription by RNA polymerase II. *Nature* **377**, 557–560 (1995).
- Serizawa, H., Conaway, J. W. & Conaway, R. C. Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. *Nature* **363**, 371–374 (1993).
- Makela, T. P. et al. A kinase-deficient transcription factor TFIIF is functional in basal and activated transcription. *Proc. Natl Acad. Sci. USA* **92**, 5174–5178 (1995).
- Cismowski, M. J., Laff, G. M., Solomon, M. J. & Reed, S. I. *KIN28* encodes a C-terminal domain kinase that controls mRNA transcription in *Saccharomyces cerevisiae* but lacks cyclin-dependent kinase-activating kinase (CAK) activity. *Mol. Cell. Biol.* **15**, 2983–2992 (1995).
- Valay, J.-G. et al. The *KIN28* gene is required both for RNA polymerase II mediated transcription and phosphorylation of the Rpb1p CTD. *J. Mol. Biol.* **249**, 535–544 (1995).
- Thompson, C. M. & Young, R. A. General requirement for RNA polymerase II holoenzymes *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 4587–4590 (1995).
- Thiele, D. J. *ACE1* regulates expression of the *Saccharomyces cerevisiae* metallothionein gene. *Mol. Cell. Biol.* **8**, 2745–2752 (1988).
- Qiu, H. et al. The *Saccharomyces cerevisiae* DNA repair gene *RAD25* is required for transcription by

- RNA polymerase II. *Genes Dev.* **7**, 2161–2171 (1993).
- Giardina, C. & Lis, J. T. Dynamic protein-DNA architecture of a yeast heat shock promoter. *Mol. Cell. Biol.* **15**, 2737–2744 (1995).
- Boorstein, W. R. & Craig, E. A. Structure and regulation of the *SSA4* HSP70 gene of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 18912–18921 (1990).
- Mogtaderi, Z. et al. TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* **383**, 188–191 (1996).
- Dubois, M.-F. et al. Heat-shock inactivation of the TFIIF-associated kinase and change in the phosphorylation sites of the C-terminal domain of RNA polymerase II. *Nucleic Acids Res.* **25**, 694–700 (1997).
- Lee, J. M. & Greenleaf, A. L. CTD kinase large subunit is encoded by *CTK1*, a gene required for normal growth of *Saccharomyces cerevisiae*. *Gene Expr.* **1**, 149–167 (1991).
- Liao, S.-M. et al. A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* **374**, 193–196 (1995).
- Barberis, A. et al. Contact with a component of the polymerase II holoenzyme is sufficient for gene activation. *Cell* **81**, 359–368 (1995).
- Ptashe, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
- Hengartner, C. J. et al. Association of an activator with an RNA polymerase II holoenzyme. *Genes Dev.* **9**, 897–910 (1995).
- Gonzalez-Couto, E., Klages, N. & Strubin, M. Synergistic and promoter-selective activation of transcription by recruitment of transcription factors TFIID and TFIIF. *Proc. Natl Acad. Sci. USA* **94**, 8036–8041 (1997).
- Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361–1375 (1993).
- Kim, Y.-J. et al. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
- Svestrup, J. Q. et al. Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl Acad. Sci. USA* **94**, 6075–6078 (1997).
- Rougvie, A. E. & Lis, J. T. The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804 (1988).
- Hansen, S. K. et al. Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* **91**, 71–83 (1997).
- Shen, W.-C. & Green, M. R. Yeast TAFII145 functions as a core promoter selectivity factor, not a general coactivator. *Cell* **90**, 615–624 (1997).
- Elion, E. A. & Warner, J. R. An RNA polymerase I enhancer in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**, 2089–2097 (1986).

Acknowledgements. We thank A. Bensaude, G. Faye, M. Cismowski, S. Reed, A. Greenleaf, S. Prakash and R. Young for strains; M. Cismowski for pSF-KIN28 plasmid; Z. Mogtaderi and K. Struhl for strains and plasmids used for generating copper knockout strains; S. Walker and M. Green for sequences of S1 probes; D. Thiele for advice on *CUP1* induction; C. Wang for generating the *Kin28* copper knockout strain; J. Roberts and E. Alani for critically reading the manuscript; and members of the Lis lab for criticism and encouragement. This work was supported by an NIH grant to J.T.L.

Correspondence and requests for materials should be addressed to J.T.L. (e-mail: jtl10@cornell.edu).

Atomic structure of progesterone complexed with its receptor

Shawn P. Williams & Paul B. Sigler

The Department of Molecular Biophysics and Biochemistry, and the Howard Hughes Medical Institute, Yale University, 160 Whitney Avenue, New Haven, Connecticut 06510, USA

The physiological effects of progestins are mediated by the progesterone receptor, a member of the steroid/nuclear receptor superfamily¹. As progesterone is required for maintenance of pregnancy, its receptor has been a target for pharmaceuticals². Here we report the 1.8 Å crystal structure of a progesterone-bound ligand-binding domain of the human progesterone receptor. The nature of this structure explains the receptor's selective affinity for progestins and establishes a common mode of recognition of 3-oxy steroids by the cognate receptors. Although the overall fold of the progesterone receptor is similar to that found in related receptors^{3–6}, the progesterone receptor has a quite different mode of dimerization^{3,6}. A hormone-induced stabilization of the carboxy-terminal secondary structure of the ligand-binding domain of the progesterone receptor accounts for the stereochemistry of this distinctive dimer, explains the receptor's characteristic pattern of ligand-dependent protease resistance and its loss of repression^{7,8}, and indicates how the anti-progestin RU486 might work in birth control. The structure also indicates that the analogous 3-keto-steroid receptors may have a similar mechanism of action.

Although the ligand-binding domain (LBD) of the progesterone receptor (PR) shares only 15% average pairwise sequence identity with previously solved nuclear receptors (Fig. 1a)^{3–5}, the principal

features of the fold of the nuclear receptors' ligand-binding domain are conserved in the PR and in the oestrogen receptor (ER) (Fig. 1), and hence in the entire nuclear/steroid receptor superfamily¹. The liganded PR LBD contains 10 α -helices, arranged in the distinctive 'helical sandwich'³, but contains no helix 2, and helices 10 and 11 are contiguous. Helix 12, the substructure considered to be most dependent on ligand binding⁴, spans all the layers enclosing the hormone-binding pocket (Fig. 1b, c). The bound hormone contacts residues from helices 3, 5, 7, 11 and 12, and the β -turn.

The PR LBD contains some features that are different from those of previously solved liganded receptor LBDs. The PR LBD contains a longer helix 12 (it contains 15 residues, whereas the retinoic acid receptor contains 8, the thyroid receptor contains 6, and the ER contains 9). Unlike the ER but like the other 3-keto-steroid receptors, the PR LBD has a 12-residue C-terminal extension

(residues 922–933) that is essential for hormone binding in the PR, glucocorticoid receptor (GR), and androgen receptor (AR)^{7,9–11}. This extension is tightly fixed in position by an antiparallel β -sheet interaction between amino acids 925–927 and 829–831, and impinges on the dimer interface seen in the 9-*cis* retinoic acid receptor (RXR) and in the ER, consistent with the different mode of dimerization of PR LBD (see below). Results of mutagenesis show that residues in the extension also support a repressive function that is exhibited in the absence of bound agonist⁷. The structure indicates that hormone-induced derepression may be caused by the inability of the presumed repressor to bind to the C-terminal substructures that are held back tightly against the body of the LBD, a mechanism consistent with hormone-induced inaccessibility of the same substructure to proteases¹². We discuss below the RU486-induced increase in accessibility and protease sensitivity of this region¹².

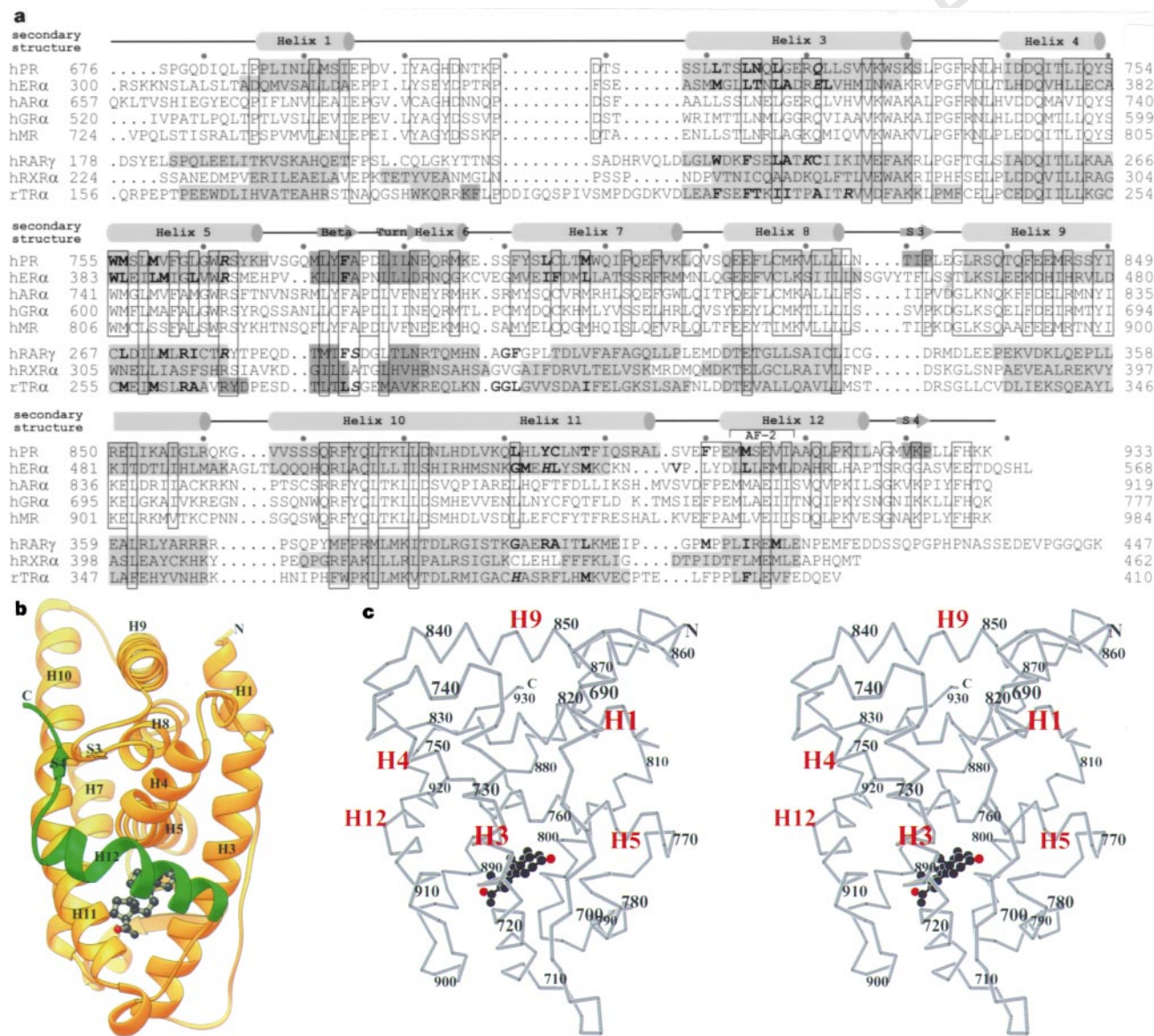


Figure 1 Comparison of the PR LBD with other LBDs. **a**, The sequence of the PR LBD is aligned with sequences of other steroid receptors (top five sequences) and of previously solved hormone-receptor LBDs (bottom three sequences) by both sequence and secondary structure with XALIGN²⁸. Secondary-structural elements of the PR LBD are indicated above the sequences and in all the receptor sequences by shading; α -helices are light grey and β -sheets are dark grey. Amino acids making van der Waals contacts with bound ligands are indicated in bold type, and amino acids making hydrogen bonds to bound ligands are bold and

italicized. Amino acids conserved among most steroid receptors or all compared receptors are boxed. The AF-2 AD core^{20,21} is indicated; h, human; r, rat; MR, mineralocorticoid receptor; RAR, retinoic acid receptor; TR, thyroid receptor. **b**, Ribbons diagram of the liganded PR LBD. H and S indicate α -helix and β -strand, respectively. Helix 12 and the C-terminal extension are green. Bound progesterone is shown as a ball-and-stick model with red oxygen atoms. **c**, Stereo view of the C α trace of the PR LBD. Bound progesterone is shown as a ball-and-stick model with red oxygen atoms.

Bound progesterone forms the core of the 'lower half' of the LBD (Fig. 1), as if it were an integral part of the tertiary structure. Specificity of hormone binding derives, in part, from the shape of the hydrophobic portion of the pocket, which complements the contours of the 18 and 19 α -methylated condensed rings, and would probably reject steroids with larger and hydrophilic α -oxygen substituents at positions 11, 18 and 19 (Fig. 2b).

The specificity determinants seen here for the 3-keto substituent of progesterone, when combined with their counterparts in the ER⁶, indicate a canonical theme for recognition of the 3-oxygen atoms in the A ring of all steroid receptors (Fig. 2). The side-chain amido NH₂ group of Gln 725 donates a hydrogen bond to the 3-keto group of progesterone. All of the steroid hormones have 3-keto groups, except for oestradiol, which has a 3-hydroxyl group, and Gln 725 is conserved in all the steroid receptors except the oestrogen receptor. In the ER, the amido group is replaced by a carboxylate oxygen of a glutamate which accepts a hydrogen bond from the unique 3-hydroxyl of oestradiol and its analogues⁶. Interestingly, this network of hydrogen bonds relies on hydrophobic hormone contacts with the walls of the steroid-binding pocket and vice versa. The required position and orientation of the Gln 725 amido group is buttressed (through an intervening fixed water site) by the backbone carbonyl of Phe 778, which in turn is fixed by a van der Waals contact with progesterone ring A. Arg 766 has a critical supporting role, underscored by its conservation at the corresponding sequence position in all steroid receptors (Fig. 2a, c). By donating a hydrogen bond (through an intervening fixed water site) to the amido oxygen of Gln 725, the guanidinium group of Arg 766 also stabilizes the

required orientation of the glutamine side chain. The intervening fixed water site also contributes a hydrogen bond to the progesterone 3-keto function, lending both specificity and rigidity to the network. The orientation of this highly flexible Arg 766 side chain is fixed, in turn, by a hydrogen bond to the carbonyl of Tyr 777, which is positioned through its rigid peptide connection to the above-mentioned Phe 778. Thus, the van der Waals contact of Phe 778 is tightly coupled through two pathways to the 3-keto-specific hydrogen-bonding system.

The specifying determinants for the methyl-ketone substituent at C17 are less clear. In the PR LBD structure, the 20-keto oxygen is too far, and is incorrectly orientated, to accept hydrogen bonds from the two nearest potential hydrogen bond donors, Asp 719 (3.8 Å distance) and Thr 894 (3.6 Å distance). Of these, Thr 894 seems the more likely discriminator as it is present only in receptors that bind hormones containing a 17-alkyl-keto group (PR, GR and the mineralocorticoid receptor), and is replaced by a large hydrophobic residue in the AR and ER, which both bind hormones with a much smaller 17-hydroxyl group.

The progesterone complex of the PR LBD crystallizes as a dyad-symmetric dimer in the asymmetric unit, with an interface that includes helix 12 (Fig. 3). The position of helix 12, and especially of the attached C-terminal extension, prevents the formation of the same dimer interface in crystalline RXR and ER^{3,6}. Instead, a much smaller and less stable¹³ dimer interface is formed that excludes only 700 Å² of solvent-accessible surface, compared with 1,703 Å² in the ER, consistent with the result that progesterone-bound PR LBD behaves as a monomer in gel-exclusion chromatography (data not

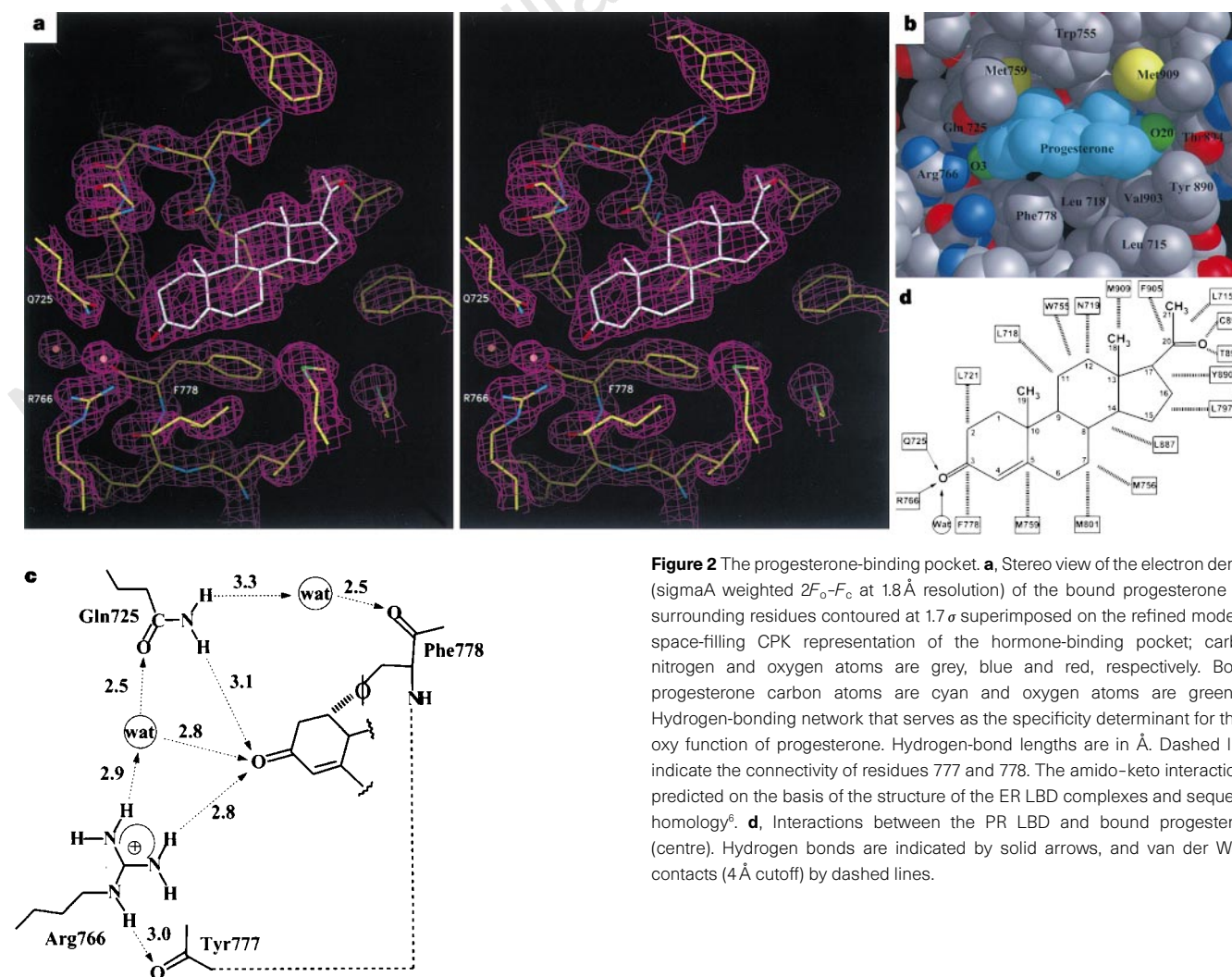


Figure 2 The progesterone-binding pocket. **a**, Stereo view of the electron density (sigmaA weighted $2F_o - F_c$ at 1.8 Å resolution) of the bound progesterone and surrounding residues contoured at 1.7 σ superimposed on the refined model. **b**, space-filling CPK representation of the hormone-binding pocket; carbon, nitrogen and oxygen atoms are grey, blue and red, respectively. Bound progesterone carbon atoms are cyan and oxygen atoms are green. **c**, Hydrogen-bonding network that serves as the specificity determinant for the 3-oxy function of progesterone. Hydrogen-bond lengths are in Å. Dashed lines indicate the connectivity of residues 777 and 778. The amido-keto interaction is predicted on the basis of the structure of the ER LBD complexes and sequence homology⁶. **d**, Interactions between the PR LBD and bound progesterone (centre). Hydrogen bonds are indicated by solid arrows, and van der Waals contacts (4 Å cutoff) by dashed lines.

shown). Although there are conflicting results concerning the requirement of progesterone for the binding of the dimeric PR to DNA^{12,14,15}, RU486, which is likely to displace helix 12 from the position seen in this structure, enhances PR dimerization and DNA binding^{16,17}, possibly by allowing the formation of an RXR- or ER-type dimer. The weakness of the progesterone-induced dimer, and inconsistent relative effects of progesterone and its antagonists¹⁸ on dimerization (in the presence and absence of DNA), indicate that further work is needed to establish the physiological relevance of the observed dimer.

We manually positioned the structure of RU486 to superimpose its condensed rings on the bound hormone so that the antagonist's O3 and C18 methyl groups overlapped those of progesterone. The *N,N*-dimethyl-aminophenyl group attached to C12 in RU486 collides severely with helix 12 and with Trp 755, which contacts helix 12. There are limitations to modelling, but these clashes would probably force a displacement of helix 12 and the C-terminal extension, accounting for the protease sensitivity¹² and increased accessibility of the PR to a presumed repressor^{7,8}. The oestrogen antagonist raloxifene also displaces helix 12 in the ER⁶, indicating a common mechanism among many antagonists of the nuclear/steroid receptors. RU486 fails to bind to the chicken and hamster PR¹⁹, probably because the bulkier side chain of cysteine, which replaces Gly 722, would sterically eliminate the binding pocket for the *N,N*-dimethyl-aminophenyl group.

Structural and mutagenesis studies of the retinoid- and thyroid-hormone receptors indicate that AF-2, a conserved amino-acid sequence that precedes and includes helix 12 (refs 20, 21) (Fig. 1a), may be a component of the interface with co-activators^{22–25}. Although the AF-2 core is conserved in the PR, helix 12 is longer, uses different interactions to stabilize its position, and distributes its functional groups differently on its outward-facing surface in the PR. Interestingly, mutations in the C-terminal extension, including deletions up to and including helix 12, allows the RU486-bound PR to activate transcription^{7,26}, as if the antirepressive effects on the C-

terminal extension were at least as important to transcriptional activation as the contribution of helix 12 to the co-activator interface. Further studies are needed to explain the diverse roles of this substructure in transcriptional control by members of the steroid receptor subfamily of the nuclear/steroid receptors. □

Methods

Expression and purification. A clone of human PR LBD was provided by B. O'Malley and M.-J. Tsai. A trypsin-limit-digest product in the presence of progesterone (amino acids 677–933) was expressed as a glutathione-S-transferase (GST)-fusion protein in *Escherichia coli* (BL21). Cells induced overnight with 25 μ M isopropyl- β -D-thiogalactoside at 15 °C in 2XYT medium supplemented with 100 nM progesterone produced 2–5 mg soluble protein per litre of culture. The soluble protein (~5% of the total expression) was extracted from the bacterial pellet with 2 M urea in a lysis buffer containing 50 mM Tris pH 8.0, 150 mM KCl, 10 mM dithiothreitol (DTT), 0.5 mM EDTA and 10 μ M progesterone, and was loaded on to a glutathione-Sepharose column. Simultaneous 2–0 M urea and 0–10% glycerol gradients in lysis buffer were applied to the protein still bound to the column. This was followed by elution with 10 mM glutathione, concentration of the eluted protein to 1 mg ml⁻¹ and overnight digestion with thrombin at 1 μ g ml⁻¹ to sever the GST. The digested protein was applied to a Mono-S column and eluted by a 50–150 mM NaCl gradient in a buffer containing 10 mM HEPES, pH 7.5, 10% glycerol, 5 mM DTT, 0.1% β -octylglucoside and 10 nM progesterone. The PR LBD eluted at 140 mM NaCl and was concentrated to 6 mg ml⁻¹.

Crystallization, structural determination and refinement. Crystals of PR LBD were obtained by vapour diffusion from drops containing a 2:1 (v/v) ratio of protein stock solution to well solution. Well solution contained 50 mM PIPES, pH 6.5, 10% glycerol, 350 mM Li₂SO₄ and 2% PEG4000. Crystals formed in 6–8 weeks at 4 °C in the space group *P*₂₁ (*a* = 58.1 Å, *b* = 64.4 Å, *c* = 69.9 Å, β = 95.7°), with two LBDs in the asymmetric unit. Derivatives were obtained by soaking crystals for 24 h in stabilizer (well solution with 5% PEG4000 and 21% ethylene glycol) containing either 1 mM K₂O₈O₄ or 0.1 mM mercuric acetate at 4 °C. All diffraction data were processed using DENZO²⁷ and SCALEPACK. A map, phased by multiple-wavelength anomalous diffraction and multiple isomorphous replacement, was improved by real-space direct methods. The PR LBD was built into this 2.6 Å experimental map using the program TurboFrodo and refined without non-crystallographic symmetry (NCS) restraints to an *R*-factor of 0.190 and *R*_{free} factor of 0.228. The two NCS-related LBDs of the asymmetric unit are related by a rotation of 178° and an axial translation of 0.8 Å. The average difference between corresponding C α positions in NCS-related domains is 0.36 Å. Phasing and refinement are shown in Table 1 in Supplementary information.

Received 2 December 1997; accepted 16 March 1998.

- Evans, R. M. The steroid and thyroid hormone receptor superfamily. *Science* **240**, 889–895 (1988).
- Baulieu, E.-E. Contraception and other clinical applications of RU486, an antiprogesterone at the receptor. *Science* **245**, 1351–1357 (1989).
- Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H. & Moras, D. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR- α . *Nature* **375**, 377–382 (1995).
- Renaud, J. P. *et al.* Crystal structure of the RAR- γ ligand-binding domain bound to all-trans retinoic acid. *Nature* **378**, 681–689 (1995).
- Wagner, R. L. *et al.* A structural role for hormone in the thyroid hormone receptor. *Nature* **378**, 690–697 (1995).
- Brzozowski, A. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–757 (1997).
- Xu, J., Nawaz, Z., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. The extreme C terminus of progesterone receptor contains a transcriptional repressor domain that functions through a putative corepressor. *Proc. Natl Acad. Sci. USA* **93**, 12195–12199 (1996).
- Klotzbucher, M., Schwerk, C., Holewa, B. & Klein-Hitpass, L. Activation of transcription by progesterone receptor involves derepression of activation functions by a cofactor. *Mol. Endocrinol.* **11**, 768–778 (1997).
- Jenster, G. *et al.* Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Mol. Endocrinol.* **5**, 1396–1404 (1991).
- Lanz, R. & Rusconi, S. A conserved carboxy-terminal subdomain is important for ligand interpretation and transactivation by nuclear receptors. *Endocrinology* **135**, 2183–2195 (1996).
- Zhang, S., Liang, X. & Danielsen, M. Role of the C terminus of the glucocorticoid receptor in hormone binding and agonist/antagonist discrimination. *Mol. Endocrinol.* **10**, 24–34 (1996).
- Allan, G. F. *et al.* Hormone and antihormone induce distinct conformational changes which are central to steroid receptor activation. *J. Biol. Chem.* **267**, 19513–19520 (1992).
- Skafar, D. F. Differential DNA binding by calf uterine estrogen and progesterone receptors results from differences in oligomeric states. *Biochemistry* **30**, 6148–6154 (1991).
- Rodriguez, R., Weigel, N. L., O'Malley, B. W. & Schrader, W. T. Dimerization of the chicken progesterone receptor in vitro can occur in the absence of hormone and DNA. *Mol. Endocrinol.* **4**, 1782–1790 (1990).
- Tetal, M. J. *et al.* Hinge and amino-terminal sequences contribute to solution dimerization of human

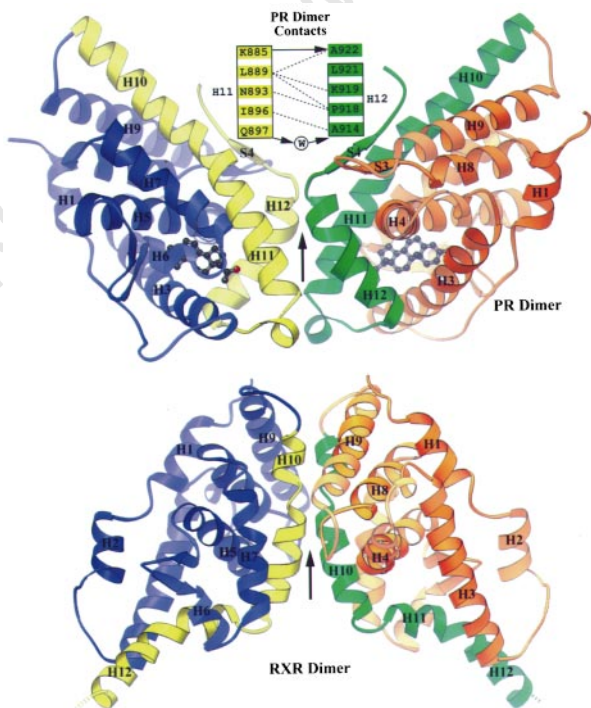


Figure 3 Comparison of the PR LBD and RXR LBD dimers. Dashed ends of RXR indicate truncation of the C-terminal five amino acids. Residues in the PR LBD that form the interface are shown at the top. Dashed lines at the top indicate van der Waals contacts (4 Å cutoff), and arrows indicate hydrogen bonds. There is also a van der Waals contact between residues Ala 922 and Leu 921.

- progesterone receptor. *Mol. Endocrinol.* **11**, 1114–1128 (1997).
16. DeMarzo, A. M., Onate, S. A., Nordeen, S. K. & Edwards, D. P. Effects of the steroid antagonist RU486 on dimerization of the human progesterone receptor. *Biochemistry* **31**, 10491–10501 (1992).
 17. Skafar, D. F. Dimerization of the RU486-bound calf uterine progesterone receptor. *J. Steroid Biochem. Mol. Biol.* **44**, 39–43 (1993).
 18. Edwards, D. P. *et al.* Progesterone receptor and the mechanism of action of progesterone antagonists. *J. Steroid Biochem. Mol. Biol.* **53**, 449–458 (1995).
 19. Benhamou, B. *et al.* A single amino acid that determines the sensitivity of progesterone receptors to RU486. *Science* **255**, 206–209 (1992).
 20. Zenke, M., Munoz, A., Sap, J., Vennstrom, B. & Beug, H. v-erbA oncogene activation entails the loss of hormone-dependent regulator activity of c-erbA. *Cell* **61**, 1035–1049 (1990).
 21. Durand, B. *et al.* Activation function 2 (AF-2) of retinoic acid receptor and 9-*cis* retinoic acid receptor: presence of a conserved autonomous constitutive activating domain and influence of the nature of the response element on AF-2 activity. *EMBO J.* **13**, 5370–5382 (1994).
 22. LeDourarin, B. *et al.* Ligand-dependent interaction of nuclear receptors with potential transcriptional intermediary factors (mediators). *Phil. Trans. R. Soc. Lond. B* **351**, 569–578 (1996).
 23. Thenot, S., Henriquet, C., Rochefort, H. & Cavaillès, V. Differential interaction of nuclear receptors with the putative human transcriptional coactivator hTIF1. *J. Biol. Chem.* **272**, 12062–12068 (1997).
 24. Collingwood, T. *et al.* A natural transactivation mutation in the thyroid hormone beta receptor: impaired interaction with putative transcriptional mediators. *Proc. Natl. Acad. Sci. USA* **94**, 248–253 (1997).
 25. Jayakumar, M., Tanen, M. & Bagchi, M. Analysis of the functional role of steroid receptor coactivator-1 in ligand-induced transactivation by thyroid hormone receptor. *Mol. Endocrinol.* **11**, 755–767 (1997).
 26. Vegeto, E. *et al.* The mechanism of RU486 antagonism is dependent on the conformation of the carboxy-terminal tail of the human progesterone receptor. *Cell* **69**, 703–713 (1992).
 27. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
 28. Wishart, D., Boyko, R. & Sykes, B. Constrained multiple sequence alignment using XALIGN. *Cabios* **10**, 687–688 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank M.-J. Tsai and B. O'Malley for the cDNA fragment containing the coding segment for the protein used here; C. Ogata (X4A-NSLS), L. Berman (X25-NSLS) and S. Ealick (CHESS) for access to and help with synchrotron radiation; R. Fletterick and D. Moras for useful structural information; D. Tanenbaum and Y. Wang for access to and discussions about their crystal structure of the oestradiol-ER-LBD complex; and members of P.B.S.'s laboratory for help with data collection. This work was supported in part by an NIH grant.

Correspondence and requests for materials should be addressed to P.B.S. (e-mail: sigler@csb.yale.edu). Coordinates have been deposited in the PDB under accession number 1a28.

corrections

Low nitrate: phosphate ratios in the global ocean

T. Tyrrell & C. S. Law

Nature **387**, 793–796 (1997)

See the Scientific Correspondence contribution on page 318 of this issue in which—owing to the discovery of errors in the international database on which this paper was based—the authors retract their main new conclusions. □

A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD

Masato Enari, Hideki Sakahira, Hideki Yokoyama, Katsuya Okawa, Akihiro Iwamatsu & Shigekazu Nagata

Nature **391**, 43–50 (1998)

We have noticed that the sequence of murine CAD (Fig. 5d) carries an error. The sequence of the amino acids from 47 to 76 should read SRLCLYEDGTVEVTDDCFGLPNDALLLL, instead of FPAVPVR-RWHGGDGRLLPGFPPTTLSSYCF as published. The open reading frame of the cDNA (pEF-mCAD) consists of 344 amino acids instead of 345 amino acids, and the mature murine CAD is a protein of 342 amino acids with a calculated relative molecular mass of 39,214.18. The sequence for mouse CAD in the DDBJ/GenBank/EMBL database (accession number AB009377) has been revised (24 February 1998). This mistake does not change the substance of our findings or the interpretation of our data. □

Phase-mapping of periodically domain-inverted LiNbO_3 with coherent X-rays

Z. W. Hu, P. A. Thomas, A. Snigirev, I. Snigireva, A. Souvorov, P. G. R. Smith, G. W. Ross & S. Teat

Nature **392**, 690–693 (1998)

The name of the first author on this Letter was incorrectly published as Z. H. Hu. □

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986